Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

Departamento de Estadística e Investigación Operativa

PhD Thesis

Hierarchical and spline-based models in space-time disease mapping

Author:
Aritz Adin Urtasun

Supervisors:
Dr. María Dolores Ugarte Martínez
Dr. Tomás Goicoa Mangado

Pamplona - Iruñea, abril 2017

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Dr. MARÍA DOLORES UGARTE MARTÍNEZ, Catedrática de Universidad adscrita al Departamento de Estadística e Investigación Operativa de la Universidad Pública de Navarra, y Dr. TOMÁS GOICOA MANGADO, Profesor Titular del Departamento de Estadística e Investigación Operativa de la Universidad Pública de Navarra,

INFORMAN

Que la presente tesis doctoral, “Hierarchical and spline-based models in space-time disease mapping” elaborada por D. ARITZ ADIN URTASUN, ha sido realizada bajo su dirección, y cumple las condiciones exigidas por la legislación vigente para optar al grado de Doctor.

Y para que así conste, firman los presentes en Pamplona, a 12 de abril de 2017

Fdo. Dra María Dolores Ugarte Martínez y Dr Tomás Goicoa Mangado

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Nire gurasoei eta nire ondoan egon diren guztiei

upna
Universidad
Pública de Navarra
Nafarroako
Unibertsitate Publikoa

Acknowledgments

I would like to thank all the people who has help me during this *journey*.

First of all, I would like to express my sincere gratitude to my supervisors, Dr. Lola Ugarte and Dr. Tomás Goicoa for their continuous support, help and guidance since my arrival at the Public University of Navarre until the final stage of this dissertation.

I am very grateful to the Department of Statistics and O.R. of the Public University of Navarre, and in particular to Dr. Ana F. Militino and Dr. Jaione Etxeberria for their help and support during these years.

I wish to thank Dr. Duncan Lee for his kindness and hospitality during my research stay at the School of Mathematics and Statistics at the University of Glasgow.

I express my gratitude to both the Spanish Ministry of Economy and Competitiveness (projects MTM2011-22664 co-funded by FEDER grants, and MTM2014-51992-R) and the Health Department of the Navarre Government (project 113, Res.2186/2014) for the financial support, and the National Epidemiology Center (area of Environmental Epidemiology and Cancer), the Basque Country Cancer Registry, and the Navarre Cancer Registry for providing the different data sets analyzed in this dissertation.

Finally, my warmest thanks to all my family, especially to my parents and my brother, and to my friends.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Contents

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

List of Figures

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

List of Tables

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Introduction

Disease mapping is an area of research of great interest in epidemiology and public health. The great variability inherent to classical risk estimation measures, such as the standardized mortality ratios (or crude rates), makes it necessary to use statistical techniques to stabilize those ratios. Many statistical models have been developed during the last years to study the geographical distribution of a disease and its evolution in time. However, the availability of high quality data recorded for many years and regions, and the emergence of new and increasingly sophisticated models, has brought new difficulties (large computing time and identifiability issues among others) that need to be thoroughly investigated. This dissertation is mainly developed within a fully Bayesian approach with the following main objectives.

The first goal is to provide a brief review of the literature in spatio-temporal disease mapping that is relevant to the research objectives of this dissertation. Some comments on the main statistical software used for model fitting and inference have been also included. In Chapter 1, the non-parametric model proposed by Knorr-Held (2000) is described in detail. Identifiability issues related to these models are also revised, and constraints to fix these problems are clearly stated. Chapter 1 also provides some insight into a new technique for Bayesian inference based on integrated nested Laplace approximations (Rue et al., 2009). The technique is known as INLA and it provides reliable results in short computational time. It can be used in the free statistical software `R` through the `R-INLA` package. Some of its most useful tools are also described.

The second objective of this dissertation is to compare some of the existing models in the literature analyzing their smoothing effects (in both space and time), and evaluating their ability to detect high-risk areas. Five different spatio-temporal models used in disease mapping have been compared in Chapter 2: the non-parametric models described by Knorr-Held (2000), a CAR-based model in space and autore-

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

gressive model in time proposed by Martínez-Beneito et al. (2008), a moving average model in space and an autoregressive model in time presented in Botella-Rocamora (2010), and three-dimensional P-spline models for spatio-temporal count data proposed by Ugarte et al. (2010b) and Ugarte et al. (2012a). To make the different terms of the models comparable, a decomposition of the estimated log-risks is computed by defining posterior spatial, temporal, and spatio-temporal "patterns". The results are illustrated using male brain cancer mortality data in Spanish provinces during the period 1986-2010. A simulation study is also conducted to compare the performance of the models in terms of sensitivity (ability to detect true high-risk regions) and specificity (ability to discard false high-risk regions) in two different scenarios (one based on the results obtained from a CAR model and a model-free scenario).

The third objective is to solve identifiability issues in spatio-temporal disease mapping models. This is a transversal objective which will be be addressed throughout the whole dissertation. Particular attention is paid on the new model proposals of Chapter 3 and Chapter 4, where the necessary set of constraints are derived by reparameterizing the random effects using the spectral decomposition of their precision matrices (Goicoa et al., 2017).

The fourth objective is to propose a new family of spatio-temporal models including two-level spatial random effects, and allowing to model the spatial and spatio-temporal effects at different administrative aggregation levels (as for example, municipalities within provinces or health areas, or counties that are grouped in states affected by similar health policies). In Chapter 3 these new model proposals are described and presented as natural extensions of the spatio-temporal CAR models (Knorr-Held, 2000). Brain cancer mortality data in the municipalities of Navarre and Basque Country during the period 1986-2008 is used to illustrate the results. In addition, a simulation study based on the analyzed municipality data is conducted to compare the single-level and two-level models in terms of smoothing and high-risk area detection. An appendix with both the identifiability constraints and the `R` code to fit these models with INLA has been included at the end of this chapter.

The fifth objective of this dissertation is to propose B-spline models in a fully Bayesian approach accounting for both the spatial and temporal correlation. In Chapter 4, different possibilities of modelling the space-time interaction using (penalized or un-penalized) B-splines are proposed. If the interest relies on analyzing the temporal evolution of risks in small areas, fitting one-dimensional temporal B-splines for each area could be preferable. If the focus is on studying the evolution in time of the geographical distribution of the risks, two-dimensional spatial B-splines for each time point can be considered. Three-dimensional P-spline models in a fully Bayesian setting are also described, in contrast to Ugarte et al. (2010b) and Ugarte

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

et al. (2012a) that used these models in an empirical Bayes disease mapping setting. Breast cancer mortality data in continental Spain during the period 1990-2010 and a simulated data for the municipalities of Navarre and Basque Country are used to illustrate the results. An appendix with both the identifiability constraints and the `R` code to fit these models with INLA has also been included at the end of this chapter.

Finally, this dissertation ends with some conclusions and comments on further research lines.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

1 Spatio-temporal disease mapping

1.1 Introduction

Models to describe the geographical distribution of a disease and its evolution in time are abundant in the spatio-temporal disease mapping literature. There has been a tremendous growth of statistical techniques for disease mapping in the last few years, mainly due to the availability of information from modern registers with high quality data recorded throughout many years and regions. The information provided by these analyses is crucial for health researchers as it helps to formulate hypothesis about the etiology of a disease and the main risk factors. Detecting hotspot areas is also of great interest for policy makers to plan effective prevention/intervention programmes.

The great variability inherent to classical risk estimation measures such as the standardized mortality ratio (SMR), makes necessary to use models that borrow strength from spatial and temporal neighbors in spatio-temporal disease mapping studies (see for example Ugarte et al., 2014). Research into spatial and spatio-temporal disease mapping has been carried out within a hierarchical Bayesian framework, with generalized linear mixed models (GLMM) playing a major role. Two main approaches have been followed for model fitting and inference, the empirical Bayes (EB) and the fully Bayes (FB) approach. Both approaches have been used in the literature and both have advantages and disadvantages (see Goicoa et al., 2012 for some discussion), but the FB approach has experienced an enormous expansion due to the advent of modern computers and free software to run Markov chain Monte Carlo (McMC) algorithms such as WinBUGS (Spiegelhalter et al., 2003), OpenBUGS (Lunn et al., 2009), JAGS (Plummer, 2016) and the new initiative STAN (Stan Development Team, 2016).

The FB approach provides posterior marginal distributions of the target parameters instead of a single point estimate. However, the posterior distributions are not

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

usually available in closed form and McMC algorithms have to be used, a computer intensive simulation-based technique. Even though these methods are very general and can be applied to virtually any model providing exact inference, in practice these algorithms can lead to high Monte Carlo errors and large computation time due to the complexity of disease mapping models (Schrödle et al., 2011) and the high dimension of the data. Moreover, specific algorithms not implemented in available software are often needed (Schmid and Held, 2004). Hence, a trade-off between exact inference, model complexity, and computing time has to be achieved. Additionally, the choice of priors for the hyperparameters is important to obtain reliable inference (see for example Wakefield, 2007; Fong et al., 2010 for some discussion). In addition to McMC, an approximate method for Bayesian inference in latent Gaussian models has recently been developed by Rue et al. (2009). The method uses integrated nested Laplace approximations (INLA) and numerical integration to estimate the posterior marginal distributions of the quantities of interest. Many latent Gaussian models admit conditional independence properties leading to sparse precision matrices, and INLA takes advantage of this to speed computation. This allows to make Bayesian inference without running long and complex McMC algorithms.

Model fitting and inference in the EB approach commonly rely on the well known penalized quasi-likelihood (PQL) technique. The maximum likelihood estimation of GLMM with counts usually requires numerical integration and PQL can reduce the problem to a series of weighted least squares regressions using a Laplace approximation to the quasi-likelihood (see Breslow and Clayton, 1993). Hence, it has been used in disease mapping as an alternative to McMC methods. It provides good point estimates for Poisson models (Dean et al., 2004), it is computationally simple and fast, and it has few convergence problems. However, it can be less accurate for binomial data, and inference relies on asymptotic distributions without clear guidelines about when this theory provides accurate inference (see Breslow, 2004 and the references therein for an in depth discussion about PQL). An additional drawback of PQL is that the variability due to the estimation of the variance components is not taken into account in the global computation of the risk variability, but some authors (see for example Ugarte et al., 2008) have developed a mean squared error estimator to avoid this limitation.

The literature about Bayesian spatio-temporal disease mapping is extensive. For example, Bernardinelli et al. (1995) use a spatio-temporal model with linear trend while Assunção et al. (2001) consider a second-degree polynomial trend model. Regarding non-parametric models, the work by Knorr-Held (2000) proposing four types of space-time interactions deserves attention. Martínez-Beneito et al. (2008) focus on an autoregressive approach to spatio-temporal disease mapping, and Ugarte et al. (2009a) compare the performance of different space-time models. Most of the research in disease mapping is based on conditional autoregressive priors (CAR) for

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

both spatial and temporal effects, extending the seminal work of Besag et al. (1991). However, other approaches based on splines have been also developed. Within an EB approach, MacNab and Dean (2001) consider autoregressive local smoothing in space and B-spline smoothing for time. Ugarte et al. (2010b, 2012b) consider a pure interaction P-spline model for space and time, and Ugarte et al. (2012a) use an ANOVA type P-spline model to describe spatio-temporal patterns of prostate cancer mortality in Spain. From a FB approach, spline smoothing has also been used in disease mapping (see for example, MacNab, 2007; MacNab and Gustafson, 2007).

This dissertation is developed within a fully Bayesian approach, using the INLA estimation technique to obtain reliable results in short computational time. This technique can be easily used in the free statistical software `R` (R Core Team, 2017) using the `R-INLA` package. Most of the work in spatial and spatio-temporal disease mapping with INLA considers the Besag et al. (1991) model (hereafter BYM model) which includes two spatial random effects: one assuming a Gaussian exchangeable prior to model unstructured heterogeneity and another one assuming an intrinsic conditional autoregressive prior (iCAR) for the spatially structured variability. See for example, Schrödle et al. (2011); Schrödle and Held (2011a); Held et al. (2010); Schrödle and Held (2011b) or Blangiardo et al. (2013). However, the iCAR prior is improper and has the undesirable large-scale property of tending to a negative pairwise correlation for regions located further apart (see MacNab, 2011; Botella-Rocamora et al., 2013). In addition, the variance components in the BYM convolution model are not identifiable from the data (MacNab, 2014) and informative hyperpriors are needed for posterior inference. In this dissertation, the prior proposed by Leroux et al. (1999) is considered to model the spatial effect. This prior has been shown to outperform the iCAR prior (Lee, 2011). The model can be implemented in `R` using the `R-INLA` package as shown in Ugarte et al. (2014).

The rest of the chapter is organized as follows. In Section 1.2, the non-parametric model proposed by Knorr-Held (2000) is described in detail, where four types of space-time interactions can be considered to model area-specific temporal evolutions. The necessary set of identifiability constraints for each model are clearly established, which are derived by reparameterizing the random effects using the spectral decomposition of their precision matrices (Goicoa et al., 2017). In Section 1.3, the INLA estimation techniques for model fitting and inference is briefly described. Finally, the `R-INLA` package and some of its more useful tools are described in Section 1.4.

1.2 Spatio-temporal models for disease mapping

Suppose that the region under study is divided into n small areas labelled as $i = 1, \ldots, n$. For each area i, data are available for different time periods labelled by

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

$t = 1, \ldots, T$. Then, conditional on the relative risk r_{it}, the number of counts O_{it} is assumed to be Poisson distributed with mean $\mu_{it} = e_{it} r_{it}$, where e_{it} is the number of expected cases for area i and time t. That is

$$\begin{aligned} O_{it}|r_{it} &\sim Poisson(\mu_{it} = e_{it} r_{it}), \\ \log \mu_{it} &= \log e_{it} + \log r_{it}. \end{aligned}$$

Here, $\log e_{it}$ is an offset and depending on the specification of $\log r_{it}$ different models are defined.

To compute the number of expected cases e_{it}, both direct and indirect 'age-and-sex' standardization procedures can be performed (note that other standardization variables could also be used in addition to age or sex). The direct method uses a single standard population to compute the 'age-and-sex' adjusted rates for each area and time period, producing rates that these areas would have if they had the same age and sex distribution as the standard population. On the other hand, the indirect method uses the same 'age-and-sex' rates (generally those computed using the information from all the areas together along the whole study period) applied to the observed population in each small area and time point. The indirect standardization procedure has been considered in all the real data analyses presented in this dissertation, so that e_{it} is computed as

$$e_{it} = \sum_{j=1}^{J} N_{itj} \frac{O_j}{N_j} \qquad i = 1, ..., n; \quad t = 1, ..., T,$$

where O_j and N_j are respectively the number of counts and the population size in 'age-and-sex' group $j \in \{1, \ldots, J\}$, so that

$$O_j = \sum_{i=1}^{n} \sum_{t=1}^{T} O_{itj} \quad \text{and} \quad N_j = \sum_{i=1}^{n} \sum_{t=1}^{T} N_{itj}.$$

Then, e_{it} represents the number of cases we would expect if the area i in time point t behaves as the whole region during the studied period.

A wide range of spatio-temporal models for disease mapping has been proposed in the literature, most of them based on CAR models extending the well known BYM model (Besag et al., 1991). Probably, the parametric model with linear time trend proposed by Bernardinelli et al. (1995) and the non-parametric models including different types of space-time interactions described in Knorr-Held (2000) are the most used models in space-time disease mapping. In the parametric model proposed

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

by Bernardinelli et al. (1995), the log-risks can be modelled as

$$\log r_{it} = \eta + \xi_i + (\beta + \varphi_i) \cdot t \tag{1.1}$$

where η is an intercept, ξ_i is the spatial effect, β represents an overall linear time trend and φ_i measures the deviation of area i from the global trend (also called differential time trend). A modification of this model was used by Ugarte et al. (2014). These authors use the Leroux et al. (1999) CAR prior distribution (LCAR) for the spatial effect ξ_i instead of the originally proposed iCAR prior. That is, the prior for the spatial random effects $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)'$ is given by

$$\boldsymbol{\xi} \sim N(\mathbf{0}, [\tau_\xi(\lambda_\xi \mathbf{R}_\xi + (1 - \lambda_\xi)\mathbf{I}_n)]^{-1}), \tag{1.2}$$

where λ_ξ is a spatial smoothing parameter taking values between 0 and 1, $\mathbf{I}_n$ is an identity matrix of dimension $n \times n$, and $\mathbf{R}_\xi$ is the $n \times n$ spatial neighborhood matrix with diagonal elements equal to the number of neighbors of each area and non-diagonal elements $(\mathbf{R}_\xi)_{ij} = -1$ if areas i and j are neighbors and $(\mathbf{R}_\xi)_{ij} = 0$ otherwise. Here, two areas are considered as neighbors if they share a common border. Note that when $\lambda_\xi = 0$ the LCAR prior reduces to an exchangeable prior $\boldsymbol{\xi} \sim N(\mathbf{0}, \tau_\xi^{-1}\mathbf{I}_n)$, whereas the iCAR prior $\boldsymbol{\xi} \sim N(\mathbf{0}, [\tau_\xi \mathbf{R}_\xi]^-)$ is obtained when $\lambda_\xi = 1$. The symbol $^-$ denotes the Moore-Penrose generalized inverse of a matrix. For the differential trend $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_n)'$, an exchangeable distribution $\boldsymbol{\varphi} \sim N(0, \tau_\varphi^{-1}\mathbf{I}_n)$ or an iCAR prior distribution $\boldsymbol{\varphi} \sim N(0, [\tau_\varphi \mathbf{R}_\xi]^-)$ can be considered.

However, the assumption of a linear time trend may be very unrealistic in practice, where it is common to observe more general temporal trends due to improvement in treatments, screening and early detection programmes, and research advances in general. A natural extension to Equation (1.1) is to drop out linearity and assume non-parametric trends. Slight modifications of the models proposed by Knorr-Held (2000) were considered by Ugarte et al. (2014). There, the log-risks are modelled as

$$\log r_{it} = \eta + \xi_i + \phi_t + \gamma_t + \delta_{it} \tag{1.3}$$

where η quantifies the logarithm of the global risk, ξ_i is the spatial component, ϕ_t and γ_t represent the unstructured and structured temporal effects respectively, and δ_{it} is the space-time interaction effect. Note that dropping the interaction terms leads to additive models. All the components in Equation (1.4) can be modelled as Gaussian Markov random fields (GMRF) (see Rue and Held, 2005), and prior densities can be written according to some structure matrices. Again, the LCAR prior was considered for the spatial random effect $\boldsymbol{\xi}$. The unstructured temporal random effects ϕ_t were modelled as independent and identically distributed normal random variables with mean 0 and precision τ_ϕ. That is, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_T)' \sim N(\mathbf{0}, \tau_\phi^{-1}\mathbf{I}_T)$, where $\mathbf{I}_t$ is the $T \times T$ identity matrix. For the structured temporal

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Table 1.1: Specification and rank deficiency for the four possible types of space-time interaction proposed by Knorr-Held (2000).

Space-time interaction	$\mathbf{R}_\delta$	Rank deficiency of $\mathbf{R}_\delta$	
		RW1 prior for $\boldsymbol{\gamma}$	RW2 prior for $\boldsymbol{\gamma}$
Type I	$\mathbf{I}_n \otimes \mathbf{I}_T$	–	–
Type II	$\mathbf{I}_n \otimes \mathbf{R}_\gamma$	n	$2n$
Type III	$\mathbf{R}_\xi \otimes \mathbf{I}_T$	T	T
Type IV	$\mathbf{R}_\xi \otimes \mathbf{R}_\gamma$	$n+T-1$	$2n+T-2$

random effects $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_T)'$, random walks of first (RW1) or second order (RW2) prior distributions were assumed, i.e., $\boldsymbol{\gamma} \sim N(\mathbf{0}, [\tau_\gamma \mathbf{R}_\gamma]^-)$, where $\mathbf{R}_\gamma$ is the $T \times T$ structure matrix of a RW1/RW2.

The interaction random effect $\boldsymbol{\delta} = (\delta_{11}, \ldots, \delta_{1T}, \ldots, \delta_{n1}, \ldots, \delta_{nT})'$ was assumed to follow the multivariate normal distribution $\boldsymbol{\delta} \sim N(\mathbf{0}, [\tau_\delta \mathbf{R}_\delta]^-)$, where $\mathbf{R}_\delta$ is the $nT \times nT$ matrix obtained as the Kronecker product of the corresponding spatial and temporal structure matrices (see Clayton, 1996), where four types of interactions can be considered. In Type I interactions ($\mathbf{R}_\delta = \mathbf{I}_n \otimes \mathbf{I}_T$), all parameters δ_{it} are a priori independent without any structure in space and time. In Type II interactions ($\mathbf{R}_\delta = \mathbf{I}_n \otimes \mathbf{R}_\gamma$), each $\delta_{i\cdot}$ for $i = 1, \ldots, n$, follows a random walk independent from all other regions; i.e., temporal trends are different from region to region, and do not have any structure in space. In Type III interactions ($\mathbf{R}_\delta = \mathbf{R}_\xi \otimes \mathbf{I}_T$), each $\delta_{\cdot t}$ for $t = 1, \ldots, T$, follows an independent iCAR prior distribution; i.e., different spatial distributions for each time point without any temporal structure are assumed. Finally, in Type IV interactions ($\mathbf{R}_\delta = \mathbf{R}_\xi \otimes \mathbf{R}_\gamma$), each δ_{it} is completely dependent over space and time. That is, different temporal trends are assumed from region to region, but trends from neighboring regions tend to be similar. The structure matrices for the different type of interactions and their rank deficiencies are summarized in Table 1.1.

In practice, the temporal effect of the data is usually structured, so the uncorrelated temporal component ϕ_t can be removed and the following model is considered

$$\log r_{it} = \eta + \xi_i + \gamma_t + \delta_{it}. \tag{1.4}$$

In this model, identifiability problems arise because the overall level can be absorbed by both the spatial and time effects, and the interaction terms are confounded with the main effects. To ensure model identifiability, sum-to-zero constraints are usually imposed over the random effects of the model (see for example Knorr-Held, 2000; Schmid and Held, 2004 or Schrödle et al., 2011). Necessary identifiability constraints using RW1 or RW2 prior for the temporally structured random effect and different

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

types of space-time interactions are summarized in Table 1.2. The details of how these sum-to-zero constraints solve the identifiability problems in spatio-temporal models are given in Goicoa et al. (2017). In this paper, the spatial, temporal, and spatio-temporal interaction random effects are reparameterized using the spectral decompositions of their precision matrices to establish the appropriate identifiability constraints, removing the combinations of the random effects that are in the span of the fixed effects (Reich et al., 2006; Hodges and Reich, 2010).

This procedure is briefly described in what follows. Let us consider a Gaussian random effect $\mathbf{a} \sim N(\mathbf{0}, [\tau\mathbf{Q}]^{-})$, with precision matrix $\tau\mathbf{Q}$. The spectral decomposition of $\mathbf{Q}$ is given by

$$\mathbf{Q} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}' = [\mathbf{U}_r : \mathbf{U}_s] \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \widetilde{\boldsymbol{\Sigma}} \end{pmatrix} \begin{bmatrix} \mathbf{U}_r' \\ \mathbf{U}_s' \end{bmatrix},$$

where $\mathbf{U} = [\mathbf{U}_r : \mathbf{U}_s]$ is an orthogonal matrix whose columns, $\mathbf{U}_r$ and $\mathbf{U}_s$, are the eigenvectors of $\mathbf{Q}$ having null and non-null eigenvalues respectively, and $\widetilde{\boldsymbol{\Sigma}}$ is a diagonal matrix with the non-null eigenvalues of $\mathbf{Q}$ in the main diagonal. Then, as $\mathbf{U}$ is orthogonal,

$$\mathbf{a} = \mathbf{U}\mathbf{U}'\mathbf{a} = [\mathbf{U}_r : \mathbf{U}_s] \begin{bmatrix} \mathbf{U}_r' \\ \mathbf{U}_s' \end{bmatrix} \mathbf{a}.$$

So, the random effect $\mathbf{a}$ can be reparameterized as $\mathbf{a} = \mathbf{X}\boldsymbol{\beta}_a + \mathbf{Z}\boldsymbol{\alpha}_a$, where

$$\begin{aligned} \mathbf{X} &= \mathbf{U}_r, \quad & \boldsymbol{\beta}_a &= \mathbf{U}_r'\mathbf{a}, \\ \mathbf{Z} &= \mathbf{U}_s, \quad & \boldsymbol{\alpha}_a &= \mathbf{U}_s'\mathbf{a}, \end{aligned} \qquad \text{and} \qquad \boldsymbol{\alpha}_a \sim N(\mathbf{0}, [\tau\widetilde{\boldsymbol{\Sigma}}]^{-1}). \tag{1.5}$$

Deleting the repeated columns obtained in the design matrices of the spatial, temporal, and spatio-temporal random effects of Equation (1.4) circumvents the identifiability issues, which implies suitable sum-to-zero constraints (see Goicoa et al., 2017 for details). The procedure will be used to derive the necessary constraints for the model proposals in Chapter 3 and Chapter 4.

1.3 Model fitting and inference

Model fitting and inference in spatio-temporal disease mapping models have usually been done using either an empirical Bayes (EB) or fully Bayes (FB) approach. In the EB approach, penalized quasi-likelihood (PQL) has been widely used (see for example MacNab and Dean, 2001; Dean et al., 2004; Ugarte et al., 2008, 2009b, 2010b). From a FB perspective, usually Markov chain Monte Carlo (McMC) techniques have been used because the posterior distributions cannot be obtained in

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Table 1.2: Identifiability constraints for the spatio-temporal CAR models described in Equation (1.4).

Interaction	RW1 prior for $\boldsymbol{\gamma}$	RW2 prior for $\boldsymbol{\gamma}$
Type I	$\sum_{i=1}^{n}\xi_i = 0, \quad \sum_{t=1}^{T}\gamma_t = 0,$ $\sum_{i=1}^{n}\sum_{t=1}^{T}\delta_{it} = 0$	$\sum_{i=1}^{n}\xi_i = 0, \quad \sum_{t=1}^{T}\gamma_t = 0,$ $\sum_{i=1}^{n}\sum_{t=1}^{T}\delta_{it} = \sum_{i=1}^{n}\sum_{t=1}^{T}t\delta_{it} = 0$
Type II	$\sum_{i=1}^{n}\xi_i = 0, \quad \sum_{t=1}^{T}\gamma_t = 0,$ $\sum_{t=1}^{T}\delta_{it} = 0$, for $i = 1, \ldots, n$	$\sum_{i=1}^{n}\xi_i = 0, \quad \sum_{t=1}^{T}\gamma_t = \sum_{t=1}^{T}t\gamma_t = 0,$ $\sum_{t=1}^{T}\delta_{it} = 0$, for $i = 1, \ldots, n$
Type III	$\sum_{i=1}^{n}\xi_i = 0, \quad \sum_{t=1}^{T}\gamma_t = 0,$ $\sum_{i=1}^{n}\delta_{it} = 0$, for $t = 1, \ldots, T$	$\sum_{i=1}^{n}\xi_i = 0, \quad \sum_{t=1}^{T}\gamma_t = 0,$ $\sum_{i=1}^{n}\delta_{it} = 0$, for $t = 1, \ldots, T$
Type IV	$\sum_{i=1}^{n}\xi_i = 0, \quad \sum_{t=1}^{T}\gamma_t = 0,$ $\sum_{t=1}^{T}\delta_{it} = 0$, for $i = 1, \ldots, n$ $\sum_{i=1}^{n}\delta_{it} = 0$, for $t = 1, \ldots, T$	$\sum_{i=1}^{n}\xi_i = 0, \quad \sum_{t=1}^{T}\gamma_t = 0,$ $\sum_{t=1}^{T}\delta_{it} = 0$, for $i = 1, \ldots, n$ $\sum_{i=1}^{n}\delta_{it} = 0$, for $t = 1, \ldots, T$

closed form (see for example Bernardinelli et al., 1995; Knorr-Held and Besag, 1998; Knorr-Held, 2000; Best et al., 2005; Martínez-Beneito et al., 2008 or Ugarte et al., 2009a). In the following section, the INLA methodology is shortly described, because this is the fitting technique that will be used in Chapter 3 and Chapter 4.

1.3.1 Integrated nested Laplace approximations (INLA)

The INLA approach recently proposed by Rue et al. (2009), is a deterministic algorithm for Bayesian inference based on integrated nested Laplace approximations. INLA is especially designed for *latent Gaussian models* (a subclass of structured additive regression models), which are flexible enough to be used in many differ-

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

ent types of applications. See Rue et al. (2017) for a review of recent examples of applications using the `R-INLA` package.

In these models, the response variable $\mathbf{y} = (y_1, \ldots, y_N)'$ is assumed to belong to an exponential family, where the mean μ_i is linked to a predictor ν_i trough a link function $g(\cdot)$, such that $g(\mu_i) = \nu_i$. The structure additive predictor ν_i is defined as follows

$$\nu_i = \eta + \sum_{j=1}^{J} \beta_j u_{ji} + \sum_{l=1}^{L} f_l(z_{li}) \quad \text{for } i = 1, \ldots, N. \tag{1.6}$$

where η is an intercept, the coefficients $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_J\}$ represents the linear effect of some covariates $\mathbf{u} = (u_1, \ldots, u_J)'$, and $f = \{f_1(\cdot), \ldots, f_L(\cdot)\}$ are unknown functions of the covariates $\mathbf{z} = (z_1, \ldots, z_L)'$. Note that a very flexible class of models are defined, since very different forms can be assumed for the unknown functions $f_l(\cdot)$, such as smooth and nonlinear effects of covariates, and temporal or spatial random effects among others.

The models described in Section 1.2 fit into this framework and are usually built as Bayesian hierarchical models with three stages. The first stage is the observational model $\pi(\mathbf{y}|\mathbf{x})$, where $\pi(\cdot|\cdot)$ denotes the conditional density and $\mathbf{y}$ is the vector of observations. We assume that y_i are conditionally independent given the vector of all the latent (non-observable) components of interest $\mathbf{x}$ and the vector of hyperparameters $\boldsymbol{\theta}$, so the distribution of the N observations is given by the likelihood

$$\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^{N} \pi(y_i|x_i, \boldsymbol{\theta}).$$

The second stage is the latent Gaussian field $\pi(\mathbf{x}|\boldsymbol{\theta})$, where a multivariate Gaussian prior with zero mean and precision matrix $\mathbf{Q}$ is assumed for $\mathbf{x}$. This precision matrix typically depends on the hyperparameters $\boldsymbol{\theta}$ (third stage), which are not necessarily Gaussian. That is, $\mathbf{x} \sim N(\mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta}))$ with density function given by

$$\pi(\mathbf{x}|\boldsymbol{\theta}) = (2\pi)^{-N/2} |\mathbf{Q}(\boldsymbol{\theta})|^{1/2} \exp\left(-\frac{1}{2}\mathbf{x}'\mathbf{Q}(\boldsymbol{\theta})\mathbf{x}\right).$$

The components of the latent Gaussian field $\mathbf{x}$ are supposed to be conditionally independent with the consequence that $\mathbf{Q}(\boldsymbol{\theta})$ is a sparse precision matrix (Blangiardo and Cameletti, 2015, Chapter 4.7.1). Note that if the components x_i and x_j are conditionally independent given all the other components $\mathbf{x}_{-ij}$, that is, if the joint conditional distribution can be factorized as $\pi(x_i, x_j|\mathbf{x}_{-ij}) = \pi(x_i|\mathbf{x}_{-ij})\pi(x_j|\mathbf{x}_{-ij})$, then $\mathbf{Q}_{ij}(\boldsymbol{\theta}) = 0$ and vice versa (Rue and Held, 2005, Chapter 2, Theorem 2.2). This specification is known as *latent Gaussian Markov random field* (GMRF). Therefore, numerical methods for sparse matrices can be used when making inference with

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

GMRFs, which are much quicker than general algorithms for dense matrices.

Note that in the particular disease mapping model of Equation (1.4), the latent Gaussian field is defined as $\mathbf{x} = (\eta, \boldsymbol{\xi}^{'}, \boldsymbol{\phi}^{'}, \boldsymbol{\gamma}^{'}, \boldsymbol{\delta}^{'})^{'}$, while the unknown precision parameters and the spatial smoothing parameter form the vector of hyperparameters $\boldsymbol{\theta} = (\tau_\xi, \lambda_\xi, \tau_\phi, \tau_\gamma, \tau_\delta)^{'}$.

In the following, we briefly explain the approximate Bayesian inference strategy of INLA. For further details see Rue et al. (2009) or Blangiardo and Cameletti (2015). The main goal is to estimate the marginal posterior distributions for each element of the GMRF

$$\pi(x_i|\mathbf{y}) = \int \pi(x_i, \boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} = \int \pi(x_i|\boldsymbol{\theta}, \mathbf{y}) \pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \tag{1.7}$$

and for each element of the hyperparameter vector

$$\pi(\theta_k|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}_{-k}.$$

The key feature of the INLA approach is to construct nested approximations of Equation (1.7). To do that, it is necessary to compute $\pi(\boldsymbol{\theta}|\mathbf{y})$ (and then the relevant marginals $\pi(\theta_k|\mathbf{y})$ can be obtained), and $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$, which is needed to compute the marginal posteriors $\pi(x_i|\mathbf{y})$. For the first task, the Laplace approximation method described in Tierney and Kadane (1986) can be used, so that the joint posterior density of the hyperparameters $\pi(\boldsymbol{\theta}|\mathbf{y})$ is approximated as

$$\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) \propto \frac{\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) \pi(\mathbf{x}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta})}{\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})}, \tag{1.8}$$

where the denominator $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ denotes the Gaussian approximation to the full conditional distribution of $\mathbf{x}$, and $\mathbf{x}^*(\boldsymbol{\theta})$ is the mode for a given $\boldsymbol{\theta}$. To integrate out the uncertainty with respect to $\boldsymbol{\theta}$, it is essential to explore the properties of Equation (1.8) and find good evaluation points θ_k for a numerical integration of Equation (1.7). This is done by an iterative algorithm (Rue et al., 2009). Additionally, an appropriate area weight Δ_k must be assigned to each θ_k. Details about how posterior marginals $\pi(\theta_k|\mathbf{y})$ are computed using numerical integration of an interpolant are available in Martins et al. (2013).

To compute the marginal distributions $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$, three different approaches are possible: a Gaussian approximation, a full Laplace approximation, and a simplified Laplace approximation. In the Gaussian approximation, the posterior conditional distributions $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$ are directly approximated as the marginals from $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$. This approximation is the fastest option and often gives accurate results in short computational time, but according to Rue and Martino (2007) unsatisfactory results

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

can be obtained due to errors in the location of the posterior marginals, errors due to the lack of skewness or both. The approximation can be improved through applying another Laplace approximation to $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$ similar to the one described in Equation (1.8). However, this "full Laplace" strategy can be computationally expensive. That is the reason why Rue et al. (2009) develop the simplified Laplace approximation based on a Taylor's series expansion of the full Laplace approximation. This method is less time consuming and gives accurate results in many applications.

Finally, an approximation to the posterior marginal density of Equation (1.7) is given by

$$\tilde{\pi}(x_i|\mathbf{y}) = \sum_k \tilde{\pi}(x_i|\theta_k, \mathbf{y})\tilde{\pi}(\theta_k|\mathbf{y})\Delta_k.$$

1.4 The R-INLA package

The INLA methodology is implemented in the open source *GMRFLib* library written in C and Fortran (Martino and Rue, 2009). An interface with the free statistical software R (R Core Team, 2017), called R-INLA, is also available allowing model specification and fitting within R. The package can be downloaded and installed in R by typing

```
> install.packages("INLA", repos="https://www.math.ntnu.no/inla/R/stable")
```

for the stable version, or

```
> install.packages("INLA", repos="https://www.math.ntnu.no/inla/R/testing")
```

for the testing version. To upgrade the package (type `inla.version()` to find out the currently installed version), use either the `inla.upgrade(testing=TRUE)` or `inla.upgrade(testing=FALSE)` commands. Documentation for the package, many worked examples, and a discussion forum are also available in the R-INLA website http://www.r-inla.org/.

As mentioned in the previous section, fixed effects, smooth and nonlinear terms, and random effects can be included in a `formula` argument using the `f()` function. The interface is flexible enough to allow for the specification of different latent models and prior distributions for the hyperparameters (see Section 1.4.1 and Section 1.4.3 for details). We run the INLA algorithm with a call to the `inla()` function

```
> inla(formula, family=<family>, data=<data>, ...)
```

where `formula` has been previously defined, `family` is a string indicating the likelihood family[1], and `data` is a data frame or list containing all the variables included in

[1] Type `names(inla.models()$likelihood)` to obtain the list of available likelihoods.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

the model. Many other additional arguments can be included into the `inla` function. See `help(inla)` for a complete list.

The output of the function is an object of class `inla`, a list containing all the results which can be explored with the `names()` function. By default, marginal distributions for the latent field and for the hyperparameters are computed. In addition, the marginal posterior distribution of the linear predictor can be computed using the `control.predictor=list(compute=TRUE)` argument. Other features as the integration strategy for $\pi(\theta_k|\mathbf{y})$ (`"auto"` (*default*), `"ccd"`, `"grid"` or `"eb"`) and the approximation strategy for $\pi(x_i|\boldsymbol{\theta},\mathbf{y})$ (`"gaussian"`, `"simplified.laplace"` (*default*) or `"laplace"`) can be also controlled within the `control.inla=list(...)` argument.

Many examples of regression models, area and point-level spatial and spatio-temporal processes, as well as the corresponding `R` code for model fitting in `R-INLA` can be found in Blangiardo and Cameletti (2015). In the following sections, a detailed description of some `R-INLA` (version 0.0-1480869339, dated 2016-12-04) model fitting tools described through this dissertation have been included.

1.4.1 Models for the latent Gaussian field

Many different latent models are implemented in the `R-INLA` package. The list of all available models can be obtained typing

```
> names(inla.models()$latent)
 [1] "linear"       "iid"          "mec"          "meb"
 [5] "rgeneric"     "rw1"          "rw2"          "crw2"
 [9] "seasonal"     "besag"        "besag2"       "bym"
[13] "bym2"         "besagproper"  "besagproper2" "fgn"
[17] "ar1"          "ar"           "ou"           "generic"
[21] "generic0"     "generic1"     "generic2"     "generic3"
[25] "spde"         "spde2"        "spde3"        "iid1d"
[29] "iid2d"        "iid3d"        "iid4d"        "iid5d"
[33] "2diid"        "z"            "rw2d"         "rw2diid"
[37] "slm"          "matern2d"     "copy"         "clinear"
[41] "sigm"         "revsigm"      "log1exp"      "logdist"
```

For each model, a detailed description and usage examples are provided in http://www.r-inla.org/models/latent-models. Some of them are briefly described in what follows. Assuming that $\mathbf{x} = (x_1, \ldots, x_k)'$ is a vector of length k:

- The `"iid"` model defines an *independent random Gaussian noise* (or *exchangeable*) prior for $\mathbf{x}$. That is

$$\mathbf{x} \sim N(\mathbf{0}, \tau^{-1}\mathbf{I}_k),$$

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

where τ is the precision parameter and $\mathbf{I}_k$ is the identity matrix of dimension $k \times k$. This model is specified inside the `f()` function as

```
> f(x, model="iid", ..., hyper=list(prec=list(...)))
```

- The `"besag"` model defines an *intrinsic CAR* prior for $\mathbf{x}$. That is

$$\mathbf{x} \sim N(\mathbf{0}, [\tau \mathbf{R}]^{-}),$$

where τ is the precision parameter and $\mathbf{R}$ is the $k \times k$ spatial neighborhood matrix. This model is specified inside the `f()` function as

```
> f(x, model="besag", graph=<graph>, ...,
+   hyper=list(prec=list(...)))
```

where the spatial neighborhood matrix $\mathbf{R}$ is passed to the program through the `graph` argument (an `inla.graph` object, a symmetric matrix or a filename containing the graph). By default, this model imposes the sum-to-zero constraint $\sum_{i=1}^{k} x_i = 0$ (`constr=TRUE`).

- The `"bym"` model defines the *BYM* (or *convolution*) prior for $\mathbf{x}$ proposed by Besag et al. (1991). That is

$$\mathbf{x} = \mathbf{u} + \mathbf{v}; \quad \text{with} \quad \begin{array}{l} \mathbf{u} \sim N(\mathbf{0}, [\tau_u \mathbf{R}]^{-}), \\ \mathbf{v} \sim N(\mathbf{0}, \tau_v^{-1} \mathbf{I}_k) \end{array}$$

where $\mathbf{u} = (u_1, \ldots, u_k)'$ is the spatially structured component with precision parameter τ_u (iCAR prior) and $\mathbf{v} = (v_1, \ldots, v_k)'$ represents the unstructured spatial component with precision parameter τ_v (iid prior). This model is specified inside the `f()` function as

```
> f(x, model="bym", graph=<graph>, ...,
+   hyper=list(prec.spatial=list(...), prec.unstruct=list(...)))
```

Since each data point is represented by two random effects, only their sum is identifiable. The `"bym"` model computes both the posterior distribution of $\mathbf{u} + \mathbf{v}$ (first k elements), and the posterior distribution of the spatially structured effect $\mathbf{u}$ (elements from $k+1$ to $2k$). By default, the sum-to-zero constraint $\sum_{i=1}^{k}(u_i + v_i) = 0$ (`constr=TRUE`) is fixed.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

- The `"rw1"` model defines a *first order random walk* prior for $\mathbf{x}$. It is constructed assuming independent increments

$$\Delta x_i = x_i - x_{i+1} \sim N(0, \tau^{-1}) \quad \text{for } i = 1, \ldots, k-1$$

where τ is the precision parameter. This model is specified inside the `f()` function as

```
> f(x, model="rw1", ..., hyper=list(prec=list(...)))
```

By default, the sum-to-zero constraint $\sum_{i=1}^{k} x_i = 0$ (`constr=TRUE`) is set.

- The `"rw2"` model defines a *second order random walk* prior for $\mathbf{x}$. It is constructed assuming independent second-order increments

$$\Delta^2 x_i = x_i - 2x_{i+1} + x_{i+1} \sim N(0, \tau^{-1}) \quad \text{for } i = 1, \ldots, k-2$$

where τ is the precision parameter. This model is specified inside the `f()` function as

```
> f(x, model="rw2", ..., hyper=list(prec=list(...)))
```

By default, the sum-to-zero constraint $\sum_{i=1}^{k} x_i = 0$ (`constr=TRUE`) is considered.

- The `"generic0"` model defines a generic prior for $\mathbf{x}$ such that

$$\mathbf{x} \sim N(\mathbf{0}, [\tau \mathbf{C}]^{-1}),$$

where τ is the precision parameter and $\mathbf{C}$ is a structure (symmetric) matrix of dimension $k \times k$ defined by the user. This model is specified inside the `f()` function as

```
> f(x, model="generic0", Cmatrix=<Cmat>, ...,
+   hyper=list(prec=list(...)))
```

where the structure matrix $\mathbf{C}$ is passed to the program through the `Cmat` argument (a dense or sparse-matrix).

- The `"generic3"` model defines a generic prior for $\mathbf{x}$ such that

$$\mathbf{x} \sim N\left(\mathbf{0}, \left[\sum_{i=1}^{m} \tau_i \mathbf{C}_i\right]^{-1}\right),$$

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

where τ_i is the specific precision parameter of the structure matrix $\mathbf{C}_i$ (of dimension $k \times k$) defined by the user. This model is specified inside the `f()` function as

```
> f(x, model="generic3", Cmatrix=<list.Cmat>, ...,
+   hyper=list(prec1=list(...),prec2=list(...),...))
```

where `list.Cmat` is a list of length m (maximum 10) with the $\mathbf{C}_i$ matrices.

1.4.2 Implementing the LCAR prior

As already mentioned, the Leroux et al. (1999) CAR prior distribution is considered in this dissertation for the spatial random effect. Recall that a LCAR prior distribution for $\mathbf{x} = (x_1, \ldots, x_k)'$ is given by

$$\mathbf{x} \sim N(\mathbf{0}, [\tau(\lambda \mathbf{R} + (1-\lambda)\mathbf{I}_k)]^{-1}),$$

where τ is the precision parameter, λ is the spatial smoothing parameter, $\mathbf{R}$ is the spatial neighborhood matrix of dimension $k \times k$ and $\mathbf{I}_k$ is an identity matrix.

This model was not originally available in R-INLA[2], but Ugarte et al. (2014) show how to built this prior distribution using the `"generic1"` model. According to the R-INLA documentation, this model defines the following prior for $\mathbf{x}$

$$\mathbf{x} \sim N\left(\mathbf{0}, \left[\tau(\mathbf{I}_k - \frac{\beta}{\lambda_{\max}}\mathbf{C})\right]^{-1}\right), \tag{1.9}$$

where τ is the precision parameter, $\mathbf{C}$ is a structure (symmetric) matrix of dimension $k \times k$ defined by the user, and $\lambda_{\max}$ is the maximum eigenvalue of $\mathbf{C}$, which allows β to be in the range $\beta \in [0, 1)$.

Let us define $\mathbf{C}$ as

$$\mathbf{C} = \mathbf{I}_k - \mathbf{R} = \begin{cases} -n_i + 1, & i = j \\ 1, & i \sim j \\ 0, & \text{otherwise} \end{cases} \tag{1.10}$$

where n_i is the number of neighbors of the ith area. Ugarte et al. (2014) proves that for this matrix $\lambda_{\max}$ equals 1, so the covariance matrix defined in Equation (1.9)

[2]At the present time, an experimental version of the model is implemented under the name `"besagproper2"`.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

takes the following expression

$$\left[\tau(\mathbf{I}_k - \frac{\beta}{\lambda_{\max}}\mathbf{C})\right]^{-1} = [\tau(\mathbf{I}_k - \beta(\mathbf{I}_k - \mathbf{R}))]^{-1} = [\tau(\beta\mathbf{R} + (1-\beta)\mathbf{I}_k)]^{-1},$$

which matches the parameterization of the covariance matrix of the LCAR prior with $\beta = \lambda$. So, the LCAR prior can be specified inside the `f()` function as

```
> f(x, model="generic1", Cmatrix=<C.Leroux>, constr=TRUE, ...,
+   hyper=list(prec=list(...),beta=list(...)))
```

where `C.Leroux` is the structure matrix defined in Equation (1.10).

1.4.3 Prior distribution for the hyperparameters

Similar to the latent models, several prior distributions are implemented in `R-INLA` for the hyperparameters θ_k. The list of all available priors can be obtained typing

```
> names(inla.models()$prior)
 [1] "normal"              "gaussian"        "wishart1d"
 [4] "wishart2d"           "wishart3d"       "wishart4d"
 [7] "wishart5d"           "loggamma"        "minuslogsqrtruncnormal"
[10] "logtnormal"          "logtgaussian"    "flat"
[13] "logflat"             "logiflat"        "mvnorm"
[16] "pc.ar"               "none"            "invalid"
[19] "betacorrelation"     "logitbeta"       "pc.prec"
[22] "pc.dof"              "pc.cor0"         "pc.cor1"
[25] "pc.fgnh"             "pc.spde.GA"      "pc.matern"
[28] "pc.range"            "pc"              "ref.ar"
[31] "jeffreystdf"         "expression:"     "table:"
```

See the web page http://www.r-inla.org/models/priors for a detailed description and examples of some of these priors. A novel approach using penalised complexity priors (PC priors) is described in Simpson et al. (2017).

In all the models for latent Gaussian fields described in Section 1.4.1, prior distributions for the precision parameters have to be specified. By default, log-Gamma distribution with parameters 1 and 5e-05 are given to the log-precision parameters in `R-INLA`, that is,

$$\theta = \log\tau \sim \text{logGamma}(1,5\text{e-}05).$$

Note that for the `"bym"` model, the vector of hyperparameters is represented as $\boldsymbol{\theta} = (\log\tau_u, \log\tau_v)$, where τ_u and τ_v are respectively the precision parameters of

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

the spatially structured and unstructured components of the model. For the LCAR model described in Section 1.4.2, the vector of hyperparameters is represented as $\boldsymbol{\theta} = (\log \tau, \text{logit}(\beta))$, with default hyperprior distribution

$$\text{logit}(\beta) = \log\left(\frac{\beta}{1-\beta}\right) \sim N(0, 0.1).$$

In addition to the prior distributions already implemented in R-INLA, the "table" and "expression" priors allow the user to define any possible prior not implemented yet. Instead of using the default hyperpriors given by R-INLA, more suitable priors have been implemented when analyzing real data examples in this dissertation. Specifically, an improper uniform prior distribution on the positive real line for the standard deviation, i.e., $\sigma = 1/\sqrt{\tau} \sim U(0, \infty)$; and a standard uniform distribution for the spatial smoothing parameter, i.e., $\beta \sim U(0, 1)$, have been defined for the hyperparameters of the random effects. However, INLA only allows to define an expression for the log-density $\theta_1 = \log \tau$ in the first case and $\theta_2 = \text{logit}(\beta)$ in the second case. So, appropriate transformations are necessary to obtain equivalent distributions in each case.

Note that the $\sigma \sim U(0, \infty)$ prior distribution can be translated to an equivalent distribution on the log-precision scale, by making

$$\pi(\theta_1) = \pi(\log \tau) = \pi(\sigma)\left|\frac{\partial \sigma}{\partial \log \tau}\right| \propto 1 \cdot \left|\frac{\partial \exp(-\log \tau/2)}{\partial \log \tau}\right| \propto \exp(-\log \tau/2),$$

and it can be implemented in R-INLA as

```
> sdunif = "expression:
+  logdens = -log_precision/2;
+  return(logdens)"
```

Once the "sdunif" prior distribution has been defined, it can be included inside the f() function as

```
> f(x, model=<model>, ..., hyper=list(prec=list(prior=sdunif)))
```

In a similar way, accounting for $\theta_2 = \text{logit}(\beta) = \log\left(\frac{\beta}{1-\beta}\right)$, the density function of θ_2 is expressed as

$$\pi(\theta_2) = \pi(\text{logit}(\beta)) = \pi(\beta) \cdot \left|\frac{\partial \beta}{\partial \theta_2}\right| = \pi(\beta) \cdot \frac{\exp(\theta_2)}{(1+\exp(\theta_2))^2} = \pi(\beta) \cdot \beta(1-\beta)$$

To define the standard uniform distribution $\beta \sim U(0, 1) \equiv \text{Beta}(1, 1)$, the log-density of $\pi(\theta_2)$ can be implemented in R-INLA as

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

```
> lunif = "expression:
+  a = 1;
+  b = 1;
+  beta = exp(theta)/(1+exp(theta));
+  logdens = lgamma(a+b)-lgamma(a)-lgamma(b)+
+            (a-1)*log(beta)+(b-1)*log(1-beta);
+  log_jacobian = log(beta*(1-beta));
+  return(logdens+log_jacobian)"
```

Once the `"lunif"` prior distribution has been defined, it can be included inside the `f()` function as

```
> f(x, model=<model>, ...,
+   hyper=list(beta=list(prior=lunif,initial=0)))
```

1.4.4 Posterior distribution of linear combinations

Depending on the context, it might be necessary to compute the posterior marginals for linear combinations of the elements in the latent field ('fixed' or 'random' effects) or for the linear predictor of the model. Details on how to compute these linear combinations within `R-INLA` are described in what follows.

For the first case, assume that our interest is in computing the posterior marginals of

$$\mathbf{w} = \mathbf{A}\mathbf{x},$$

where $\mathbf{x} = (x_1, \ldots, x_k)'$ is the latent field and $\mathbf{A}$ is a $p \times k$ matrix where p is the number of linear combinations of the latent field. The functions `inla.make.lincomb()` and `inla.make.lincombs()` can be used to define a linear combination or several linear combinations, respectively. As remarked in Martins et al. (2013), two different approaches are provided in `R-INLA`. The first approach creates an enlarged latent field $\tilde{\mathbf{x}} = (\mathbf{x}, \mathbf{w})$ and then posterior marginals for $\tilde{\mathbf{x}}$ are computed with the INLA method using the Gaussian, simplified Laplace or full Laplace approximation strategies described in Section 1.3.1. However, the addition of many linear combinations will lead to more dense precision matrices which will consequently slow down the computations. The second approach does not include $\mathbf{w}$ in the latent field, but performs a post-processing of the resulting output given by INLA and approximates the posterior marginals of $\mathbf{w}$ by a Gaussian distribution with

$$E[\mathbf{w}|\boldsymbol{\theta}, \mathbf{y}] = \mathbf{A}\boldsymbol{\mu}^* \quad \text{and} \quad \text{Var}[\mathbf{w}|\boldsymbol{\theta}, \mathbf{y}] = \mathbf{A}(\mathbf{Q}^*)^{-1}\mathbf{A}^T,$$

where $\boldsymbol{\mu}^*$ is the mean of the marginal approximation $\tilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y})$ and $\mathbf{Q}^*$ is the precision matrix of the Gaussian approximation $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ used in Equation (1.8). This

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

approach leads to a much faster approximation of the posterior marginals for $\mathbf{w}$ and that is why this is the default method in R-INLA. However, more accurate approximations can be obtained switching to the first approach, if necessary, by including the following argument into the inla function

```
> inla(formula, family=<family>, data=<data>, ...,
+      control.inla=list(lincomb.derived.only=FALSE))
```

Spatial, temporal and spatio-temporal "patterns" are defined in Section 2.4 as linear combinations of the linear predictor. Thus, the inla.make.lincombs() function is used to compute the posterior marginal distributions of these patterns.

However, as in the B-spline models proposed in Chapter 4, the response might depend on a linear combination of the latent field. Suppose that we want to fit the following model

$$y_i = \eta + \sum_{j=1}^{k} a_{ij} x_i, \quad \text{for} \quad i = 1, \ldots, N$$

where $\mathbf{y} = (y_1, \ldots, y_N)'$ is the response vector, η is an intercept, $\mathbf{A} = (a_{ij})$ is a design matrix of dimension $N \times k$ and $\mathbf{x} = (x_1, \ldots, x_k)'$ is a vector of unknown coefficients where an exchangeable prior is considered, i.e., $\mathbf{x} \sim N(\mathbf{0}, \tau \mathbf{I}_k)$. This model can be specified in R-INLA as

```
> # Define the formula argument
> formula <- y ~ -1 + intercept + f(x,model="iid",...)
>
> # Call to the inla() function
> eta <- rep(1,N)
> data <- list(intercept=c(1,rep(NA,k)),x=c(NA,1:k))
>
> inla(formula, family=<family>, data=data, ...,
+      control.predictor=list(A=cBind(eta,<A.matrix>),...))
```

where A.matrix contains the coefficients of the linear combinations. Internally, R-INLA adds another layer in the hierarchical model

$$\boldsymbol{\nu}^* = \mathbf{A}\boldsymbol{\nu},$$

so that the likelihood function is linked to the latent field through $\boldsymbol{\nu}^*$ instead of $\boldsymbol{\nu}$, i.e.,

$$\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^{N} \pi(y_i|\nu_i^*, \boldsymbol{\theta}).$$

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

According to Martins et al. (2013), this feature is implemented by also adding $\boldsymbol{\nu}^*$ to the latent model, where the conditional distribution for $\boldsymbol{\nu}^*$ has the mean $\mathbf{A}\boldsymbol{\nu}$ and the precision matrix $\kappa_A \mathbf{I}$, where the constant κ_A is set to a high value. In terms of output from `inla`, the vector $(\boldsymbol{\nu}^*, \boldsymbol{\nu})$ will be the linear predictor.

1.4.5 Linear constraints for the latent Gaussian fields

In the spatio-temporal disease mapping models described in Section 1.2, sum-to-zero constraints are considered to ensure model identifiability between the intercept, the main spatial and temporal effects, and the space-time interaction effect.

In `R-INLA`, sum-to-zero constraints is the default option for intrinsic models (see Section 1.4.1) for the latent Gaussian field $\mathbf{x}$, that is, $\sum_{i=1}^{k} x_i = 0$. Including this constraint (specified inside the `f()` function with the `constr=TRUE` argument), makes it possible to identify the intercept and the main spatial/temporal effect. However, as stated in Table 1.2, additional contraints must be imposed over the spatio-temporal interaction term depending on its prior distribution. Using the `extraconstr` argument, linear constrains such as $\mathbf{Ax} = \mathbf{b}$ can be specified, where the number of rows of $\mathbf{A}$ is equal to the number of constraints to impose over $\mathbf{x}$.

For example, let us consider the spatio-temporal model of Equation (1.4) where a RW1 prior is given to the temporal random effect $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_t)^{'}$, and a completely structured prior (Type IV) is considered for the interaction effect $\boldsymbol{\delta} = (\delta_{11}, \ldots, \delta_{1T}, \ldots, \delta_{n1}, \ldots, \delta_{nT})^{'}$. According to the Table 1.2, the following $n + T$ sum-to-zero constraints on the interaction term are needed

$$\sum_{t=1}^{T} \delta_{it} = 0, \quad \text{for } i = 1, \ldots, n.$$

$$\sum_{i=1}^{n} \delta_{it} = 0, \quad \text{for } t = 1, \ldots, T.$$

Note that these constraints can be written in the form $\mathbf{A}\boldsymbol{\delta} = \mathbf{0}$ with

$$\mathbf{A} = \begin{bmatrix} \mathbf{I}_n \otimes \mathbf{1}_T^{'} \\ \mathbf{1}_n^{'} \otimes \mathbf{I}_T \end{bmatrix}. \tag{1.11}$$

The constraints are specified in the `f()` function as

```
> f(ID.delta, model="generic0", Cmatrix=<Cmat>,
+   rankdef=<rankdef>, constr=TRUE, hyper=list(prec=...),
+   extraconstr=list(A=<A.constr>, e=rep(0,<n.constr>)))
```

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

where `Cmat` is the Kronecker product of the spatial and temporal structure matrices, `rankdef` is its rank deficiency, `A.constr` is the $(n+T)\times nT$ dimension matrix given in Equation (1.11) and `n.constr` is equal to the number of constraints to be imposed.

1.4.6 Model selection criteria

Criteria based on the deviance

When the interest is comparing different models in terms of model fitting and complexity, some criteria based of the deviance can be used. Given the data $\mathbf{y}$ with likelihood $\pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})$ the *Bayesian deviance* is defined as

$$D(\mathbf{x},\boldsymbol{\theta}) = -2\log\{(\pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})\} + 2\log\{f(\mathbf{y})\}. \tag{1.12}$$

The deviance of the model is a measure that takes account of the variability associated to the likelihood. Generally, the posterior mean deviance $\overline{D(\mathbf{x},\boldsymbol{\theta})}$ is considered as a measure of goodness of it due to its robustness. However more complex models will fit better the data, and consequently lower values of the mean deviance will be obtained. The Deviance Information Criterion (DIC, Spiegelhalter et al., 2002) is the most commonly used measure of model fit based on the deviance for Bayesian models. The DIC is computed as the sum of the posterior mean of the deviance (a measure of goodness of fit) and the number of effective parameters (a measure of model complexity),

$$\text{DIC} = \overline{D(\mathbf{x},\boldsymbol{\theta})} + p_D$$

where the quantity p_D is defined as the posterior mean of the deviance minus the deviance of the posterior means of the parameters of interest. Analogously to the Akaike information criterion (AIC), models with smaller DIC values provide better trade-off between model fit and complexity. To compute the DIC values, the option `control.compute=list(dic=TRUE)` inside the `inla()` function is used. The details about how these quantities are computed in `R-INLA` can be found in Rue et al. (2009). It is very important to know that INLA does not compute the saturated deviance, i.e., the second term in Equation (1.12) is not used. In addition INLA, instead of evaluating the deviance at the posterior mean of all parameters, evaluates the deviance at the posterior mean of the parameters and at the posterior mode of the hyperparameters. The reason is that the posterior marginals for some hyperparameters (especially the precisions) might be highly skewed.

A corrected version of the DIC proposed by Plummer (2008) has been also considered as model selection criterion, because it has been shown that DIC values may under-penalize complex models in disease mapping. The corrected DIC is defined

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

as

$$\text{DIC}_c = \overline{D(\mathbf{x}, \boldsymbol{\theta})} + \sum_{i=1}^{N} p_{D_i}/(1 - p_{D_i}) \tag{1.13}$$

where p_{D_i} is the contribution of observation i to the effective number of parameters (see Spiegelhalter et al., 2002).

The more recently derived Watanabe-Akaike information criterion (Watanabe, 2010), also known as Widely Applicable Information Criterion (WAIC), which is recommended by Gelman et al. (2014) over the DIC criterion, can be also computed by including the option `control.compute=list(waic=TRUE)`. WAIC is a method for estimating pointwise out-of-sample prediction accuracy from a fitted Bayesian model, and unlike DIC, is invariant to parametrization and also works for singular models.

Criteria based on the predictive distribution

Different scoring rules can be defined to compare the models in terms of their predictive performance by assigning a numerical score to each model based on their predictive distributions.

Given a set of spatio-temporal observations $\mathbf{y} = (y_{11}, \ldots, y_{nT})'$, the logarithmic score (see Gneiting and Raftery, 2007) is defined as

$$LS = -\sum_{i=1}^{n}\sum_{t=1}^{T} \log(\text{CPO}_{it}) \tag{1.14}$$

where $\text{CPO}_{it} = Pr(Y_{it} = y_{it}|\mathbf{y}_{-it})$ values (conditional predictive ordinate, Pettit, 1990) denotes the cross-validated predictive probability mass at the observed count y_{it}. The logarithmic score is asymptotically equivalent to the Akaike information criterion if the observations are independent (Stone, 1977). Models with smaller resulting scores will be better in terms of predictive performance.

The probability integral transform (PIT, Dawid, 1984) for each observation can be also computed to assess the predictive quality of a model, which is defined as

$$\text{PIT}_{it} = Pr(Y_{it} \leq y_{it}|\mathbf{y}_{it}),$$

that is, the cross-validated predictive cumulative distribution at the observed count y_{it}. If $\mathbf{y}$ comes from a continuous distribution, the PIT values have a standard uniform distribution. So the histogram of the computed PIT values can be used as a diagnostic tool. U-shaped histograms indicate underdispersed predictive distributions, while hump or inverse U-shaped histograms point towards overdispersion. However, in the case of count data the predictive distribution is discrete and the

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

PITs are no longer uniform under the hypothesis of an ideal forecast. In this case, the adjusted version of the PIT suggested by Czado et al. (2009) can be used instead, although it is not recommended for binary responses or Poisson with very few counts.

As described by Rue et al. (2009), both CPO and PIT quantities are computed in R-INLA without re-running the model by including into the `inla()` function the argument `control.compute=list(cpo=TRUE)`. Their accuracy in comparison with quantities that are obtained by McMC methods is discussed in Held et al. (2010). As noted in this paper, the approximation of the predictive measures might fail if the approximation of the latent field is not accurate enough. This is due to an insufficient exploration of the tail properties of involved densities. Hence, the full Laplace approximation might be obligatory to get reliable results. It is also possible to increase accuracy of the estimation for the tails of the marginal distributions by adding the option `control.inla=list(strategy="laplace", npoints=<h>)` to add more evaluation points instead of the default `npoints=9`.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

2

Evaluation of models for smoothing and detecting high-risk areas: a comparison of P-splines, autoregressive, and moving average models

2.1 Introduction

Several models have been proposed for smoothing risks in disease mapping, considering alternative ways of introducing both spatial and temporal dependence as well as spatio-temporal interactions. The non-parametric models proposed by Knorr-Held (2000) and described in Chapter 1, are possibly, the most widely used models in space-time disease mapping. However, other proposals have been also considered for smoothing risks. For example, Martínez-Beneito et al. (2008) propose CAR-based models in space and autoregressive models in time while Botella-Rocamora (2010) consider moving average models in space and autoregressive models in time. Lately, three dimensional P-spline models have been also used to smooth risks in space and time (Ugarte et al., 2010b, 2012a). Despite the fact that CAR and P-spline models have been compared in a spatial context (Goicoa et al., 2012), no comparisons have been performed in the spatio-temporal setting yet.

The aim of this chapter is to deal with this issue providing practitioners some guidance on choosing the most appropriate model in a particular situation. Brain cancer mortality data in Spanish provinces during the period 1986-2010 (already analyzed in Ugarte et al., 2014) will be used for illustration purposes. The main spatial, temporal and spatio-temporal patterns will be investigated. A simulation study will be performed to compare these models in different spatio-temporal scenarios in terms of bias, variability, sensitivity (ability to detect true positives, i.e., ability to detect true high risk regions), and specificity (ability to discard false pos-

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

itives, i.e, ability to discard false high risk regions). Identifiability issues have been taken into account when fitting the previous models, considering the corresponding constraints for each one. In addition, an adequate decomposition of the log-risks is proposed to compare the patterns induced by all the models.

This chapter is laid out as follows. In Section 2.2 a brief review of P-spline models in spatio-temporal disease mapping is given, while Section 2.3 describes the autoregressive and moving average models mentioned above. Computational issues about model estimation and a decomposition of the estimated log-risks are detailed in Section 2.4. Brain cancer mortality data in Spanish provinces are analyzed in Section 2.5 using all the models. Section 2.6 evaluates the capability of the models to detect true high risk areas performing a simulation study in two different spatio-temporal scenarios. Finally, Section 2.7 concludes with a discussion and some interesting guidelines.

2.2 P-splines in spatio-temporal disease mapping

Although splines have become popular in a spatio-temporal context, their use has been mainly limited to model temporal rather than spatial effects (MacNab and Dean, 2001; MacNab, 2007; MacNab and Gustafson, 2007). An anisotropic and non separable three-dimensional model is proposed in Ugarte et al. (2010b), extending the two-dimensional P-spline model proposed by Lee and Durbán (2009) to smooth risks in space and time. An ANOVA-type P-spline model was also considered in Ugarte et al. (2012a), allowing different smoothing parameters for the main spatial, temporal and interaction effects. In what follows, these three-dimensional P-spline models will be briefly described. See for example Goicoa et al. (2016) for a review of P-splines with B-splines bases in spatial and spatio-temporal disease mapping.

2.2.1 Interaction P-spline model

In the Interaction P-spline model (Ugarte et al., 2010b), the log-relative risks are modeled as a smooth function of the covariates

$$\log r_{it} = \eta + f(x_{1i}, x_{2i}, x_t), \quad \text{for} \quad \begin{array}{l} i = 1, \ldots, n, \\ t = 1, \ldots, T, \end{array} \tag{2.1}$$

where η is an intercept, x_{1i} and x_{2i} are the coordinates of the centroid of the area i (longitude and latitude respectively), x_t is the time, and $f(x_{1i}, x_{2i}, x_t)$ is an unknown smooth function that is assumed to be sufficiently well approximated using P-splines with B-splines bases (Eilers and Marx, 1996).

In matrix form

$$\log \mathbf{r} = (\mathbf{1}_{nT})\eta + \mathbf{B}\boldsymbol{\theta}^{(st)},$$

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

where $\mathbf{r} = (r_{11}, \ldots, r_{n1}, \ldots, r_{1T}, \ldots, r_{nT})'$, $\mathbf{1}_{nT}$ is a vector of ones of length nT, $\mathbf{B}$ is a three-dimensional B-spline basis, and $\boldsymbol{\theta}^{(st)}$ is the vector of coefficients. If $\mathbf{B}_1$, $\mathbf{B}_2$, and $\mathbf{B}_t$ are $n \times k_1$, $n \times k_2$, and $T \times k_t$ marginal B-splines bases for longitude, latitude and time respectively (with k_1, k_2 and k_t depending on the number of knots and the degree of the polynomials in the marginal bases), then the B-splines basis $\mathbf{B}$ is defined as the Kronecker product $\mathbf{B} = \mathbf{B}_t \otimes \mathbf{B}_s$, where $\mathbf{B}_s = \mathbf{B}_2 \square \mathbf{B}_1$ is the two-dimensional B-spline basis of dimension $n \times k_1 k_2$ obtained from the row-wise Kronecker product (Eilers et al., 2006) of the marginal bases for longitude $\mathbf{x}_1 = (x_{11}, \ldots, x_{1n})'$ and latitude $\mathbf{x}_2 = (x_{21}, \ldots, x_{2n})'$. That is,

$$\mathbf{B}_2 \square \mathbf{B}_1 = (\mathbf{B}_2 \otimes \mathbf{1}'_{k_1}) \odot (\mathbf{1}'_{k_2} \otimes \mathbf{B}_1), \tag{2.2}$$

where the symbol $\odot$ indicates the Hadamard product (elementwise multiplication of the two matrices), and $\mathbf{1}_{k_1}$ and $\mathbf{1}_{k_2}$ are vectors of ones of length k_1 and k_2 respectively.

To achieve smoothness a three-dimensional penalty is placed over the regression coefficients, expressed in terms of marginal penalties as

$$\mathbf{P}_{st} = \tau_1(\mathbf{I}_{k_t} \otimes \mathbf{I}_{k_2} \otimes \Delta'_{d_1}\Delta_{d_1}) + \tau_2(\mathbf{I}_{k_t} \otimes \Delta'_{d_2}\Delta_{d_2} \otimes \mathbf{I}_{k_1}) + \tau_t(\Delta'_{d_t}\Delta_{d_t} \otimes \mathbf{I}_{k_2} \otimes \mathbf{I}_{k_1}),$$

where $\mathbf{I}_{k_i}$ are identity matrices and Δ_{d_i} are difference matrices (usually of order 1 or 2) of dimension $k_i \times k_i$. The parameters τ_1, τ_2 and τ_t control the amount of smoothing in each direction.

Each coefficient, a component of $\boldsymbol{\theta}^{(st)}$, would be influenced by its row (longitude) and column (latitude) neighbors. As we have a penalty in time, it would be also affected by the time neighbors. However, the effect of the row, column and time neighbors is not the same as the smoothing parameters are different. A possible limitation of this model may be that the same smoothing parameters τ_1, τ_2 and τ_t used for longitude, latitude and time are also used for the space-time interaction. The ANOVA-type P-spline model briefly reviewed in the next subsection overcomes this potential limitation.

2.2.2 ANOVA-type P-spline model

In the ANOVA-type P-spline model, additive terms for space (longitude and latitude), time, and space-time interactions are included, modeling the log-relative risks as

$$\log r_{it} = \eta + f_1(x_{1i}, x_{2i}) + f_2(x_t) + f_3(x_{1i}, x_{2i}, x_t), \text{ for } \begin{matrix} i = 1, \ldots, n, \\ t = 1, \ldots, T, \end{matrix} \tag{2.3}$$

where η is an intercept, $f_1(x_{1i}, x_{2i})$ is a smooth surface constant along the time periods, $f_2(x_t)$ is a temporal smooth function common to all areas and $f_3(x_{1i}, x_{2i}, x_t)$

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

is a three-dimensional P-spline to model the space-time interaction. All the smooth functions can be approximated using B-spline bases as

$$\begin{aligned} f_1(\mathbf{x}_1, \mathbf{x}_2) &= \mathbf{B}_s \boldsymbol{\theta}^{(s)}, \\ f_2(\mathbf{x}_t) &= \mathbf{B}_t \boldsymbol{\theta}^{(t)}, \\ f_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_t) &= (\mathbf{B}_t \otimes \mathbf{B}_s) \boldsymbol{\theta}^{(st)}, \end{aligned}$$

where different smoothing parameters for each function f_i, $i = 1, \ldots, 3$ are considered, making the model more flexible than the Interaction P-spline model of Equation (2.1). The paper by Ugarte et al. (2012a) gives additional details on the use of this particular model in disease mapping. This model has been also used for forecasting mortality risks and counts by Etxeberria et al. (2015).

Both the Interaction and ANOVA-type P-spline models consider space-time interactions and the coefficients are penalized in space and time. In this sense, these models could resemble Type IV interaction models of Knorr-Held (2000), but they are more flexible in the sense that the degree of smoothing can vary in the three dimensions (longitude, latitude or time).

2.3 Autoregressive and moving average models

2.3.1 BYMar model

The model proposed by Martínez-Beneito et al. (2008) will be named BYMar because it models the random region effect using the convolution model of Besag et al. (1991) (called in short BYM) and the temporal correlation is induced through a first order autoregressive structure. The model is briefly described next.

$$\begin{aligned} \log r_{i1} &= \eta + \alpha_1 + (1-\rho^2)^{-1/2}(\theta_{i1} + \phi_{i1}), && i = 1, \ldots, n, \\ \theta_{i1} &\sim N(0, \tau_\theta^{-1}), && i = 1, \ldots, n, \\ \boldsymbol{\phi}_1 = (\phi_{11}, \ldots, \phi_{n1})' &\sim N(\mathbf{0}, [\tau_\phi \mathbf{R}_\phi]^-), && \text{(iCAR model)} \end{aligned} \tag{2.4}$$

where η quantifies the global log-risk, α_1 models the mean deviation of the risk in the first period from the mean level of all of them, $\mathbf{R}_\phi$ is the $n \times n$ spatial neighborhood matrix and ρ represents the temporal correlation introduced for $t = 2, \ldots, T$ as

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

$$\begin{aligned} \log r_{it} &= \eta + \alpha_t + \rho(\log r_{i,t-1} - \eta - \alpha_{t-1}) + \theta_{it} + \phi_{it}, \\ \theta_{it} &\sim N(0, \tau_\theta^{-1}), \\ \boldsymbol{\phi}_t = (\phi_{1t}, \ldots, \phi_{nt})' &\sim N(\mathbf{0}, [\tau_\phi \mathbf{R}_\phi]^-), \\ \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_t)' &\sim N(\mathbf{0}, [\tau_\alpha \mathbf{R}_\alpha]^-), \end{aligned} \tag{2.5}$$

where $\mathbf{R}_\alpha$ is the structure matrix of a random walk of order one. The temporal dependence introduced in Equation (2.5) makes the relative risks in each region and time point depend not only on their neighbors relative risks in the same time point, but also on the relative risks in previous periods. The dependence in each region among different time points has been defined as a first order autoregressive time series. Finally, spatial dependence has been introduced using a BYM prior.

In Equation (2.4), the term $(1-\rho^2)^{-1/2}$ is introduced in order to make the variance-covariance matrix of $\log \mathbf{r}_1$ equal to the stationary covariance matrix of the series $\{\log \mathbf{r}_t\}_{t=1}^{\infty}$. Then,

$$(\log \mathbf{r}_1, \log \mathbf{r}_2, \ldots, \log \mathbf{r}_T)' \,|\, \ldots \sim N((\eta \cdot \mathbf{1}_T + \boldsymbol{\alpha}) \otimes \mathbf{1}_n, \boldsymbol{\Lambda} \otimes \boldsymbol{\Sigma}),$$

where $(\log \mathbf{r}_1, \log \mathbf{r}_2, \ldots, \log \mathbf{r}_T)' \,|\, \ldots$ denotes the log-relative risk distribution given all the parameters in the different hierarchies of the model, $\boldsymbol{\Sigma}$ is the covariance matrix of $\theta_{\cdot t} + \phi_{\cdot t}$ for any period t and $\boldsymbol{\Lambda}$ denotes the correlation matrix of a first-order autoregressive time series of length T. In contrast to CAR models and the Knorr-Held (2000) proposal in particular, $\boldsymbol{\Lambda}$ depends now on one parameter, ρ, what makes this matrix more flexible than those used in the other proposals. With respect to prior distributions, a Unif$(0,1)$ prior was used for ρ, while Unif$(0,5)$ prior distributions were used for all the standard deviations (square root inverse of precision parameters τ_θ, τ_ψ and τ_α) involved in this model. The upper limit of this distribution was considered as a vague choice since it is referred to variables defined in a logarithmic scale (log-risks). Finally, an improper flat prior distribution was used for η. See Martínez-Beneito et al. (2008) for more details about this model.

2.3.2 STMARS model

The STMARS model is the spatio-temporal extension of the Spatial Moving Average Risk Smoothing (SMARS) model proposed in Botella-Rocamora et al. (2013). The SMARS model can be considered as an alternative to CAR-based processes. If CAR models are the equivalent in the spatial domain to the temporal integrated and autoregressive processes, then SMARS can be considered as the spatial equivalent to moving average processes in time. The spatio-temporal extension is defined as in the BYMar model using an autoregressive temporal structure of first order. In the

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

STMARS model, the log-risks are modeled as

$$\log r_{it} = \eta + \alpha_t + \omega_0 \psi_{it} + \omega_1 \sum_{i' \sim_1 i} \psi_{i't} + \ldots + \omega_m \sum_{i' \sim_m i} \psi_{i't},$$

where η represents the mean of the log-risks for the first period and α_t the differential risk of the period t with regards to the first (with $\alpha_1 = 0$). A RW1 structure is considered for these differential risks, i.e., $\alpha_t \sim N(\alpha_{t-1}, \tau_\alpha^{-1})$ for $t = 2, \ldots, T$. The set $i' \sim_k i$ denotes all the geographical units being k-th order neighbors of region i (those with the shortest path between both regions having k edges) and $\boldsymbol{\omega} = (\omega_0, \ldots, \omega_m)'$ is a vector weighting the contribution of the neighbors of differing orders of region i. The maximum order of the neighbors having an effect on the log-relative risks, m, is also a variable estimated by the model. As mentioned above, temporal dependence is achieved using a first-order autoregressive process on the latent effects $\boldsymbol{\psi}_{it}$, such that

$$\begin{aligned} \psi_{i1} &\sim N(0, \tau_\psi^{-1}), \\ \psi_{it} &\sim N(\rho \cdot \psi_{i,(t-1)}, (1-\rho^2)/\tau_\psi) \quad t = 2, \ldots, T. \end{aligned}$$

The set of k-th order neighbors defined for this model, as for the SMARS model, will form a kind of irregular disc centered on region i and a radius defined by k. In general, if any region is a k-th order neighbor of region i, it will be also a k'-th order neighbor for any $k' > k$. The zeroth-order neighbor of region i will be considered as its own region i. This makes neighbors of lower order to be more influential than those of higher order. Additional details about this model model can be seen in Botella-Rocamora (2010, Chapter 4).

A Unif$(0, 1)$ was used for ρ. Improper flat prior distributions were used for η and $\sigma_\psi = 1/\sqrt{\tau_\psi}$, in the latter case restricted to the positive real line. As the scale of $\boldsymbol{\psi}$ is controlled by τ_ψ, the restriction $\sum_{j=0}^{m} \omega_j = 0$ is imposed on $\boldsymbol{\omega}$. Thus, a flat Direchlet$(\mathbf{1}_{n+1})$ is used as prior distribution for $\boldsymbol{\omega}$. Finally, $P(m) \propto (m!)^{-1}$ was used as prior distribution for m. A deeper reasoning on the use of these priors is given in Botella-Rocamora et al. (2013).

2.4 Some aspects of model fitting and model comparisons

The models described in the previous sections have been fitted to the brain cancer mortality data in Spanish provinces. The same fitting techniques proposed by the authors have been considered here. Penalized quasi-likelihood (PQL) technique has been used to fit the three-dimensional P-spline models, the BYMar model has

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

been fitted in WinBUGS (Spiegelhalter et al., 2003), and finally an algorithm has been implemented in R to fit the STMARS model. The algorithm includes Gibbs sampling, Metropolis-Hastings, and Reversible Jump McMC.

Similar to the spatio-temporal CAR models described in Equation (1.4), the BYMar and P-spline models suffer from identifiability problems and constraints are required to achieve sensible results. Here, appropriate sum-to-zero constraints have been imposed over the random effects in the BYMar model. A different approach has been used to solve the identifiability issues in the P-spline models, since they are fitted using their mixed model representation. Under this formulation, identifiability problems are easily solved removing the repeated columns in the fixed effect matrices (Ugarte et al., 2010b, 2012a). It can be shown that removing these columns is equivalent to imposing sum-to-zero constraints over the P-spline regression coefficients (see Appendix 4A).

In addition, to make the different terms of the models comparable, a decomposition of the estimated log-risks is computed by defining the following posterior spatial, temporal and spatio-temporal "patterns"

$$\begin{aligned}
\eta^* &= \tfrac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\log r_{it},\\
\xi_i^* &= \tfrac{1}{T}\sum_{t=1}^{T}\log r_{it} - \eta^*,\\
\gamma_t^* &= \tfrac{1}{n}\sum_{i=1}^{n}\log r_{it} - \eta^*,\\
\delta_{it}^* &= \log r_{it} - \xi_i^* - \gamma_t^* - \eta^*.
\end{aligned}$$

It can be easily checked that the estimated log-risks can be decomposed as the sum of these patterns, i.e., $\log r_{it} = \eta^* + \xi_i^* + \gamma_t^* + \delta_{it}^*$. Note that if a completely structured interaction (Type IV) is considered for the CAR model of Equation (1.4), it holds that $\eta^* = \eta$, $\xi_i^* = \xi_i$, $\gamma_t^* = \gamma_t$ and $\delta_{it}^* = \delta_{it}$. In addition, since these patterns are centered at zero, we are able to decompose the total amount of variability of the overall log-risks as the sum of the spatial, temporal and spatio-temporal variabilities as follows

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(\log r_{it} - \eta^*)^2 = \frac{1}{n}\sum_{i=1}^{n}(\xi_i^*)^2 + \frac{1}{T}\sum_{t=1}^{T}(\gamma_t^*)^2 + \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(\delta_{it}^*)^2.$$

Then, the percentage of spatial, temporal and spatio-temporal variability for any data set can be computed. Posterior distributions of $\log \mathbf{r}$, and therefore of η^*, $\boldsymbol{\xi}^*$, $\boldsymbol{\gamma}^*$ and $\boldsymbol{\delta}^*$, can be easily computed if McMC methods are used for model estimation. Additionally, expressing these patterns as linear combinations of the linear predictor, posterior distributions can be also computed in INLA using the

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

`inla.make.lincombs` function (see Section 1.4.4). Regarding PQL, only point estimates of the different patterns are provided, and hence the delta method has been used to compute the corresponding variances of α^*, ξ_i^*, γ_t^* and δ_{it}^*, being able to derive pointwise confidence bands for these patterns.

Denoting $\log \hat{r}_{it} = \hat{u}_{it}$, the overall log-risk pattern α^* can be expressed as the linear function

$$\begin{array}{rcl} f_\alpha : \mathbb{R}^{nT} & \longrightarrow & \mathbb{R} \\ \hat{\mathbf{u}} & \rightsquigarrow & f_\alpha(\hat{\mathbf{u}}) = \alpha^* = \frac{1}{nT}(\hat{u}_{11} + \cdots + \hat{u}_{nT}). \end{array}$$

So, using the delta method

$$\widehat{\mathrm{Var}}(f_\alpha(\hat{\mathbf{u}})) = \widehat{\mathrm{Var}}(\alpha^*) = (\nabla f_\alpha(\hat{\mathbf{u}}))' \widehat{\mathrm{Var}}(\hat{\mathbf{u}})(\nabla f_\alpha(\hat{\mathbf{u}})),$$

where the gradient of f_α is computed as $\nabla f_\alpha(\hat{\mathbf{u}}) = (\frac{1}{nT}, \ldots, \frac{1}{nT})'$, and $\widehat{\mathrm{Var}}(\hat{\mathbf{u}})$ is the estimated covariance matrix of the log-risks. See Ainsworth and Dean (2006) or MacNab and Lin (2009) for details about computation of the covariance matrix of the estimated log-risks under the PQL technique.

In a similar way, the spatial pattern is expressed as

$$\begin{array}{rcl} f_\xi : \mathbb{R}^{nT} & \longrightarrow & \mathbb{R}^{n} \\ \hat{\mathbf{u}} & \rightsquigarrow & f_\xi(\hat{\mathbf{u}}) = \boldsymbol{\xi}^*, \end{array}$$

where the i-th component of vector $\boldsymbol{\xi}^*$ is

$$\left[-\frac{1}{nT}(\hat{u}_{11} + \cdots + \hat{u}_{nT}) + \frac{n}{nT}(\hat{u}_{i1} + \cdots + \hat{u}_{iT}) \right]_i$$

The temporal pattern is expressed as

$$\begin{array}{rcl} f_\gamma : \mathbb{R}^{nT} & \longrightarrow & \mathbb{R}^{T} \\ \hat{\mathbf{u}} & \rightsquigarrow & f_\gamma(\hat{\mathbf{u}}) = \boldsymbol{\gamma}^*, \end{array}$$

where the t-th component of vector $\boldsymbol{\gamma}^*$ is

$$\left[-\frac{1}{nT}(\hat{u}_{11} + \cdots + \hat{u}_{nT}) + \frac{T}{nT}(\hat{u}_{1t} + \cdots + \hat{u}_{nt}) \right]_t$$

And the spatio-temporal pattern is expressed as

$$\begin{array}{rcl} f_\delta : \mathbb{R}^{nT} & \longrightarrow & \mathbb{R}^{nT} \\ \hat{\mathbf{u}} & \rightsquigarrow & f_\delta(\hat{\mathbf{u}}) = \boldsymbol{\delta}^*, \end{array}$$

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

where the it-th component of vector $\boldsymbol{\delta}^*$ is

$$\left[\hat{u}_{it} + \frac{1}{nT}(\hat{u}_{11} + \cdots + \hat{u}_{nT}) - \frac{n}{nT}(\hat{u}_{i1} + \cdots + \hat{u}_{iT}) - \frac{T}{nT}(\hat{u}_{1t} + \cdots + \hat{u}_{nt})\right]_{it}$$

So, if the gradient of the functions f_ξ, f_γ, f_δ are computed, the estimated variances of these patterns using the delta method are expressed as

$$\widehat{\text{Var}}(f_\xi(\hat{\mathbf{u}})) = \widehat{\text{Var}}(\xi^*) = (\nabla f_\xi(\hat{\mathbf{u}}))' \widehat{\text{Var}}(\hat{\mathbf{u}})(\nabla f_\xi(\hat{\mathbf{u}})),$$

$$\widehat{\text{Var}}(f_\gamma(\hat{\mathbf{u}})) = \widehat{\text{Var}}(\gamma^*) = (\nabla f_\gamma(\hat{\mathbf{u}}))' \widehat{\text{Var}}(\hat{\mathbf{u}})(\nabla f_\gamma(\hat{\mathbf{u}})),$$

$$\widehat{\text{Var}}(f_\delta(\hat{\mathbf{u}})) = \widehat{\text{Var}}(\delta^*) = (\nabla f_\delta(\hat{\mathbf{u}}))' \widehat{\text{Var}}(\hat{\mathbf{u}})(\nabla f_\delta(\hat{\mathbf{u}})).$$

2.5 Illustration

We first illustrate and compare all the approaches described in this chapter by modeling the spatio-temporal evolution of male brain cancer mortality in Spanish provinces during the period 1986-2010. These data are already analyzed in Ugarte et al. (2014), fitting a total of 18 (parametric and non-parametric) CAR models using different types of space-time interactions. The neighborhood structure was defined using adjacency between areas, i.e., two provinces are considered as neighbors is they share a common border. Finally, a model including a random spatial effect with a LCAR prior, a structured temporal effect with a RW1 prior, and a completely structured interaction term (Type IV interaction, $\mathbf{R}_\delta = \mathbf{R}_\xi \otimes \mathbf{R}_\gamma$) was selected. This model will be used to compare the results obtained with the P-spline, BYMar and STMARS models.

To fit the P-spline models, cubic B-splines and second order penalties have been used for the spatial dimensions, considering eleven (equidistant) internal knots for longitude and latitude; for the temporal dimension, cubic B-splines, second order penalties and seven internal knots have been considered. When fitting the BYMar model in WinBUGS, after discarding 5000 iterations of burn-in, 10000 more were generated and only one in every ten was saved following the recommendations given by Martínez-Beneito et al. (2008). Finally, for the STMARS model 10000 burn-in iterations were discarded, generating 50000 more iterations and saving only one in every fifty. All described models were fitted on a PC with Inter(R) Core(TM) i5-2400 CPU @3.10GHz processor, using `R` version 3.2.2 (64 bits). Computational times for fitted models are displayed in Table 2.1.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Table 2.1: Approximated computational times (in seconds) for fitted models.

CAR	Interaction P-spline	ANOVA P-spline	BYMar	STMARS
460(*)	3350	4440	14800	1050

*Full Laplace approximation (`strategy="laplace"`)

Maps with the geographical patterns of brain cancer mortality risk $\zeta_i = \exp(\xi_i^*)$ estimated by each model are shown on top of Figure 2.1. Significantly high risk provinces are drawn at the bottom of Figure 2.1. The risk of each province is classified as high with respect to the whole of Spain if the lower bound of the corresponding upper one-sided confidence or credible interval is greater than one. We are computing intervals at different levels given by the scale shown at the bottom-right of Figure 2.1. For example, red coloured provinces are those whose lower bound of the 90 % upper one-sided interval is greater than one.

Little differences are observed between the P-spline models. In general, more gradual changes are observed when comparing P-splines with CAR or autoregressive models. In Figure 2.2, the temporal patterns (common to all regions) of brain cancer mortality risks, $\exp(\gamma_t^*)$, are plotted. Finally, spatio-temporal interaction patterns δ_{it}^* for some of the Spanish provinces are shown in Figure 2.3. Again, smoother pictures are observed when fitting P-spline models. In this particular example small differences are observed among the rest of the models.

2.6 Simulation study

In this section we present a simulation study to compare the performance of the models in two different scenarios. In the first scenario we simulate data based on the results of the CAR model fitted in Section 2.5, and the second scenario is a model-free scenario similar to the one used by Martínez-Beneito et al. (2008). In both scenarios, spatio-temporal risk surfaces, r_{it}, have been defined as the product of geographical, temporal and spatio-temporal terms for the $n = 50$ provinces in Spain. For each region i and time t, O_{it} counts are generated from a Poisson distribution with mean $\mu_{it} = e_{it} r_{it}$, where the expected number of cases computed for the brain cancer data set are multiplied by the scale factors (SF) 1/4.5, 1/8 and 1/20 in the first scenario, while expected values equal $E = 10$, 5, 3 and 1 for all areas have been considered in the second scenario. A total of 100 data sets have been generated for each subscenario.

To compare the accuracy of the models in terms of relative risk estimates (posterior means in the Bayesian case), average values of mean absolute relative bias

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

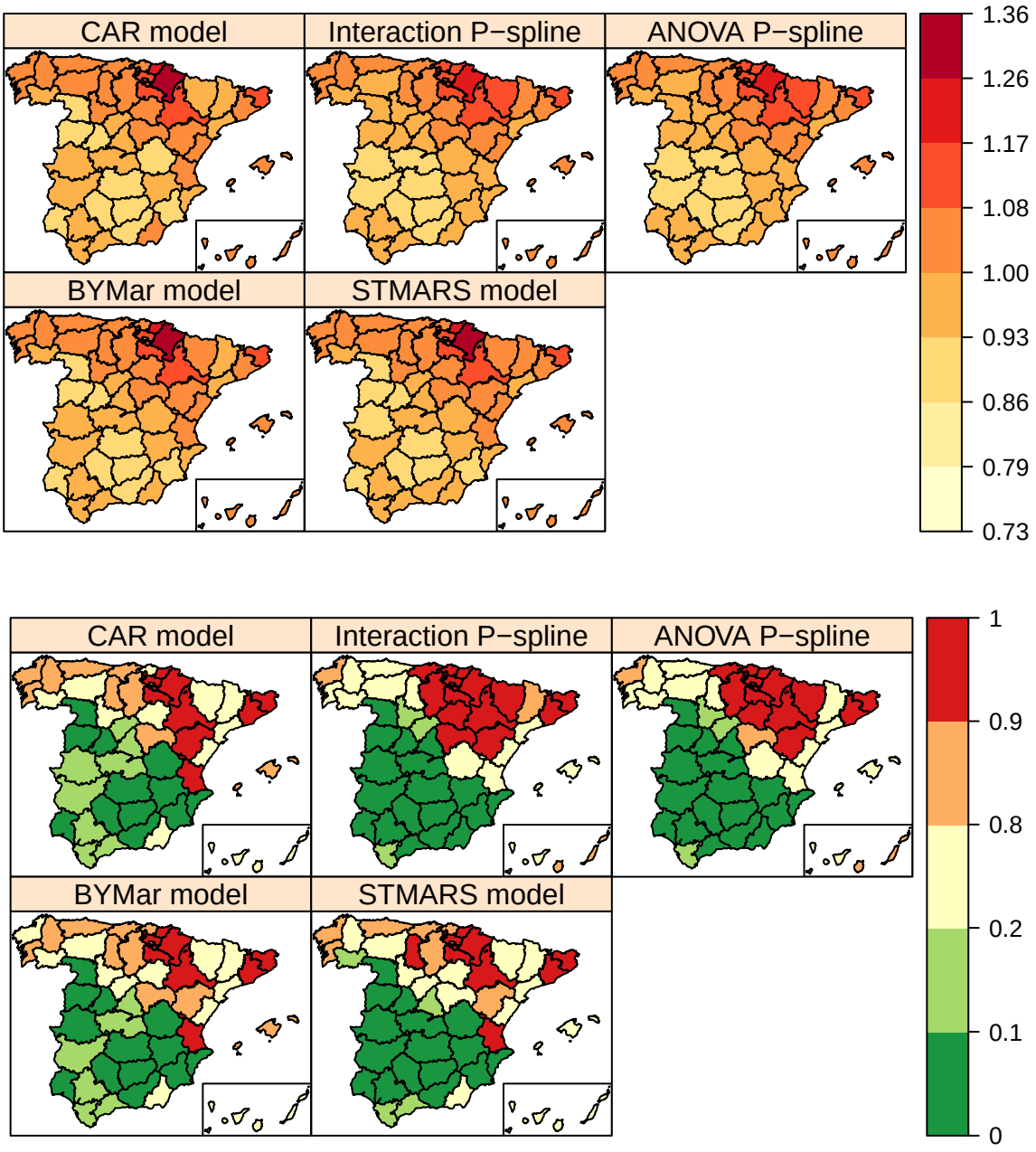


Figure 2.1: Geographical patterns of brain cancer mortality risks (top) and significantly high risk provinces (bottom).

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

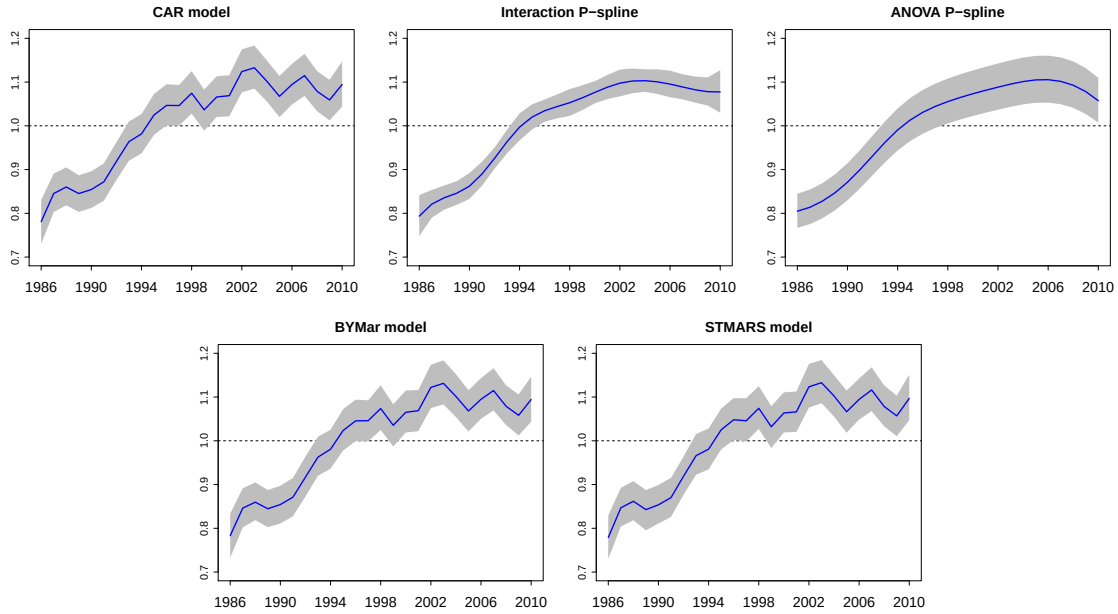


Figure 2.2: Overall temporal trends of brain cancer mortality risks. The X-axis represents years and the Y-axis gives $\exp(\gamma_t^*)$.

(MARB) and mean relative root mean prediction standard error (MRRMPSE) have been calculated for each province

$$MARB_i = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{100}\left|\sum_{k=1}^{100}\frac{\hat{r}_{it}^k - r_{it}}{r_{it}}\right|,$$

$$MRRMPSE_i = \frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{1}{100}\sum_{k=1}^{100}\left(\frac{\hat{r}_{it}^k - r_{it}}{r_{it}}\right)^2},$$

where T denotes the number of time periods (years). In our simulation study, $T = 25$ in Scenario 1 corresponding to the period 1986-2010, while $T = 15$ in Scenario 2 according to the original simulation study described in Martínez-Beneito et al. (2008).

In addition, as the simulated risks are defined as the product of geographical, temporal and spatio-temporal interaction patterns, we can also compute these measures for the previously defined ξ_i^*, γ_t^* and δ_{it}^* posterior patterns.

Upper one-sided EB confidence intervals and FB credible intervals have been computed to compare the performance of the corresponding models in terms of the ability to detect high-risk areas (sensitivity) controlling at the same time the false positive rates (1-specificity).

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

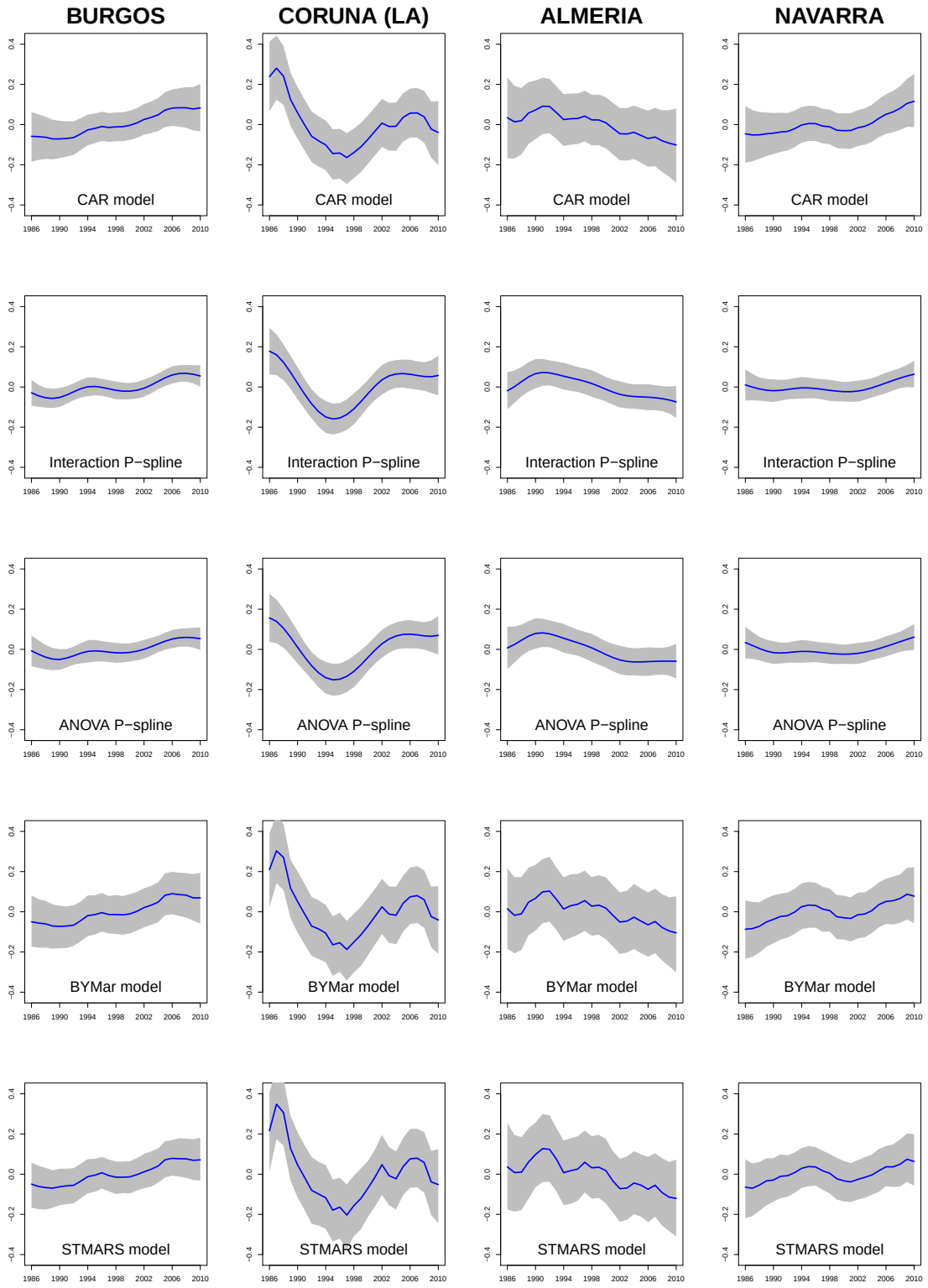


Figure 2.3: Spatio-temporal patterns δ_{it}^{*} for each province.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

2.6.1 Scenario 1

In this setting, we consider as true risk patterns those obtained from the analysis of brain cancer data using the completely structured (Type IV) spatio-temporal CAR model described in Section 2.5. The percentage of variability of the overall risk explained by each of the patterns is about 32% (spatial), 54% (temporal) and 14% (spatio-temporal). Once the 100 data sets were generated, the completely structured (Type IV interaction) CAR model and the four different models described in Section 2.2 and Section 2.3 were fitted. The same model specifications used in the illustration case have been considered, except for the ANOVA-type P-spline model, where a constrained model with equal smoothing parameters for the interaction has been fitted to avoid convergence problems.

Table 2.2 displays average values of MARB and MRRMPSE of the relative risk estimates as well as of the spatial, temporal, and spatio-temporal patterns. When the expected values match the actual scenario (SF=1), the CAR model performs, in general, the best (as expected), followed closely by BYMar model. However, when the number of expected cases starts to decrease, in particular when the number of expected cases is very small (SF=1/20), models performance change and P-spline models behave better in terms of MARB for global relative risks, spatial and temporal patterns, and in terms of MRRMPSE when estimating the spatio-temporal patterns.

True positive (TPR) and false positive (FPR) rates are shown in Table 2.3. TPR (expressed as percentages) have been computed as the proportion of high true risks ($r_{it} > 1$) that were classified as high risks. A risk is classified as a high risk if the lower bound of the corresponding upper one-sided credible/confidence interval is greater than one. FPR have been computed similarly. P-spline models show the higher percentages of TPR. Although the rate of false positives is also higher, it never overcomes 10 %. When the number of expected cases decreases, the true and false positive rates also decrease. In the case of very sparse data (SF=1/20), TPR are remarkably small for the autoregressive, CAR and STMARS models.

Finally, a comparison of the receiver-operator characteristic curves (ROC) and the area under those curves (AUC) for the different subscenarios are represented in Figure 2.4. In a ROC curve, true positive rates (y axis) are plotted against false positive rates (x axis) at different significance levels. The ROC curves enables us to compare the performance of different models in terms of sensitivity and specificity. Models with higher AUC values will be preferred in terms of classification ability (in our context, high-risk area detection). As shown in Figure 2.4, the CAR-based models and the STMARS model performs better in the scenario with higher expected values (SF=1). However, if the number of expected cases decreases, the classification ability of the ANOVA-type P-spline model outperforms the rest od the models.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Table 2.2: Scenario 1: average values of MARB and MRRMPSE for relative risks and for spatial, temporal and spatio-temporal pattern estimates based on 100 simulated data sets.

	MARB				MRRMPSE			
	r_{it}	ξ_i^*	γ_t^*	δ_{it}^*	r_{it}	ξ_i^*	γ_t^*	δ_{it}^*
SF=1								
CAR model	0.0284	0.0183	0.0078	0.0184	0.0525	0.0362	0.0190	0.0299
Interaction P-spline	0.0454	0.0329	0.0154	0.0248	0.0592	0.0425	0.0203	0.0323
ANOVA P-spline	0.0448	0.0333	0.0159	0.0230	0.0593	0.0430	0.0201	0.0322
BYMar	0.0293	0.0195	0.0078	0.0189	0.0526	0.0358	0.0190	0.0307
STMARS	0.0315	0.0224	0.0074	0.0194	0.0528	0.0362	0.0192	0.0301
SF=1/4.5								
CAR model	0.0470	0.0342	0.0141	0.0256	0.0768	0.0535	0.0316	0.0386
Interaction P-spline	0.0565	0.0404	0.0178	0.0310	0.0771	0.0550	0.0282	0.0393
ANOVA P-spline	0.0562	0.0403	0.0184	0.0275	0.0746	0.0524	0.0275	0.0399
BYMar	0.0516	0.0384	0.0153	0.0282	0.0778	0.0527	0.0318	0.0414
STMARS	0.0531	0.0404	0.0147	0.0287	0.0773	0.0527	0.0324	0.0396
SF=1/8								
CAR model	0.0524	0.0387	0.0164	0.0274	0.0850	0.0584	0.0370	0.0415
Interaction P-spline	0.0577	0.0419	0.0189	0.0317	0.0854	0.0600	0.0326	0.0421
ANOVA P-spline	0.0555	0.0428	0.0181	0.0288	0.0846	0.0589	0.0308	0.0444
BYMar	0.0607	0.0458	0.0182	0.0319	0.0858	0.0564	0.0373	0.0453
STMARS	0.0630	0.0482	0.0179	0.0331	0.0856	0.0571	0.0382	0.0426
SF=1/20								
CAR model	0.0618	0.0452	0.0221	0.0307	0.1049	0.0685	0.0516	0.0476
Interaction P-spline	0.0601	0.0424	0.0187	0.0343	0.1108	0.0806	0.0423	0.0533
ANOVA P-spline	0.0610	0.0407	0.0210	0.0334	0.1055	0.0807	0.0434	0.0448
BYMar	0.0728	0.0538	0.0263	0.0383	0.1035	0.0619	0.0520	0.0519
STMARS	0.0751	0.0560	0.0263	0.0387	0.1023	0.0625	0.0529	0.0460

2.6.2 Scenario 2

A similar scenario to that used in the simulation study presented in Martínez-Beneito et al. (2008) is considered now. Instead of simulating from a particular model, four non overlapping and dispersed clusters are defined centered in the provinces of Navarra (A), Valencia (B), Huelva (C) and La Coruña (D), with different time evolution along the study period. The maps with the temporal evolution of these clusters are shown in Figure 2.5.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Table 2.3: Scenario 1: true and false positives rates using lower one-sided credibility/confidence intervals.

	True positives			False positives		
	95% CI	90% CI	80% CI	95% CI	90% CI	80% CI
SF=1						
CAR model	0.420	0.533	0.676	0.002	0.009	0.029
Interaction P-spline	0.669	0.728	0.793	0.047	0.067	0.099
ANOVA P-spline	0.653	0.714	0.782	0.044	0.064	0.097
BYMar	0.405	0.518	0.661	0.002	0.007	0.027
STMARS	0.454	0.562	0.693	0.005	0.015	0.040
SF=1/4.5						
CAR model	0.173	0.305	0.507	0.002	0.010	0.041
Interaction P-spline	0.532	0.605	0.691	0.041	0.065	0.107
ANOVA P-spline	0.498	0.583	0.686	0.027	0.049	0.093
BYMar	0.112	0.238	0.461	0.001	0.008	0.039
STMARS	0.187	0.334	0.542	0.007	0.023	0.064
SF=1/8						
CAR model	0.111	0.224	0.430	0.002	0.008	0.039
Interaction P-spline	0.429	0.527	0.637	0.031	0.054	0.095
ANOVA P-spline	0.390	0.495	0.627	0.023	0.048	0.088
BYMar	0.044	0.131	0.349	0.001	0.007	0.038
STMARS	0.092	0.219	0.463	0.006	0.021	0.075
SF=1/20						
CAR model	0.035	0.091	0.253	0.003	0.008	0.037
Interaction P-spline	0.227	0.331	0.492	0.024	0.047	0.100
ANOVA P-spline	0.207	0.312	0.474	0.015	0.026	0.065
BYMar	0.007	0.032	0.152	0.001	0.005	0.029
STMARS	0.024	0.081	0.259	0.005	0.017	0.066

High-risk areas located on cluster A (middle-north cluster) in the first period are "moved" to cluster B (north-east cluster) during the simulated years. Although cluster C (south-west) retains its position, relative risks increase during the first half and decrease during the second half of the study period. Finally, cluster D (north-west) is gradually expanded to their neighboring areas. The percentage of variability of the overall risk explained by the interaction term is about 21%, while most of the variability is explained by the spatial pattern (about 69 % of the total variability). Unlike Scenario 1, the temporal pattern only explains about 10% of the

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

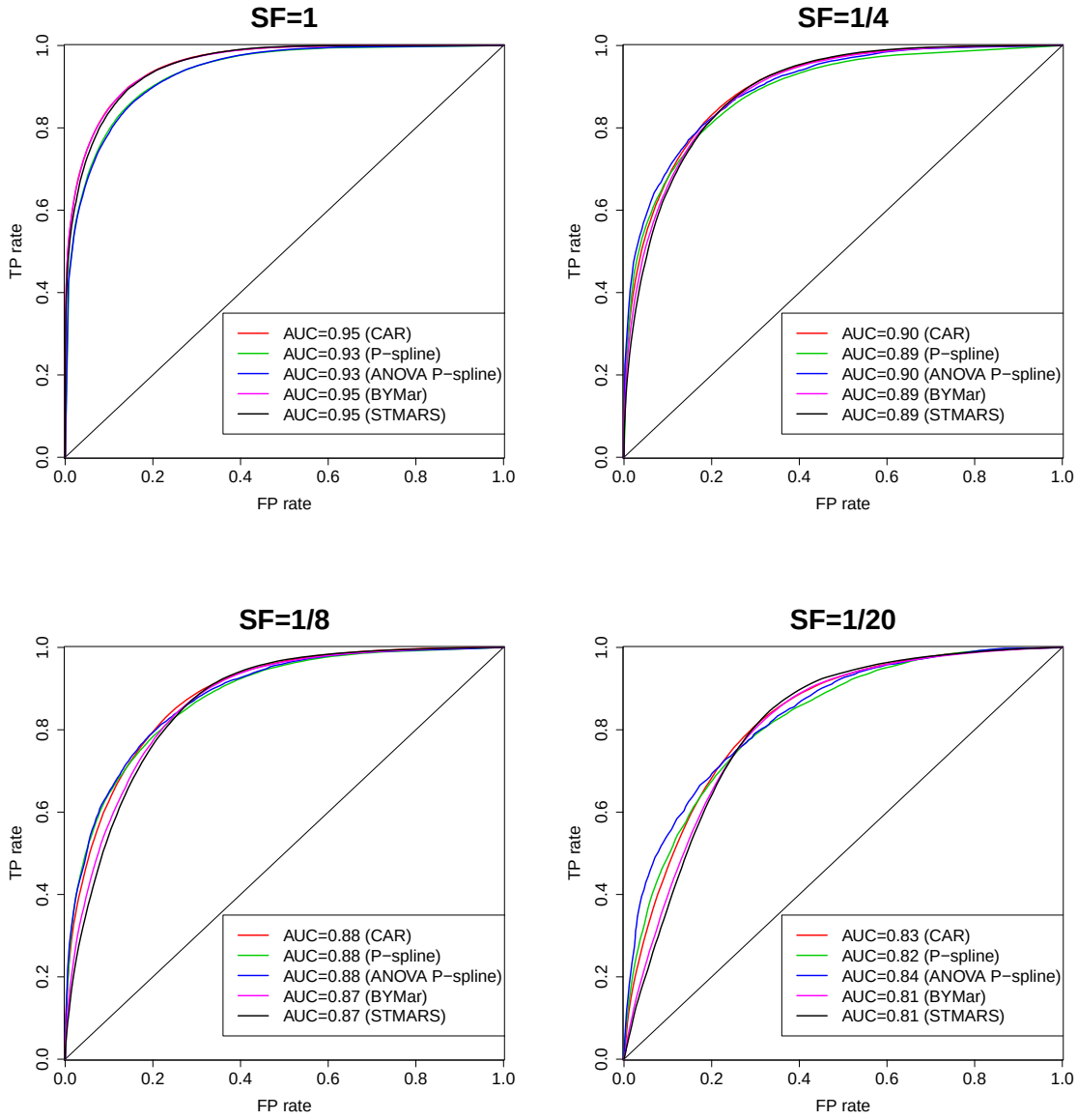


Figure 2.4: Scenario 1. AUC for different scale factor (SF) values.

total variability.

The CAR models considered in Ugarte et al. (2014) were fitted and once again, a model including a RW1 temporal effect and a Type IV interaction was selected as the best model in terms of DIC. The rest of the models have been implemented as in Scenario 1, except the ANOVA-type P-spline model, where the original model with six different smoothing parameters have been fitted. Average values of MARB and

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

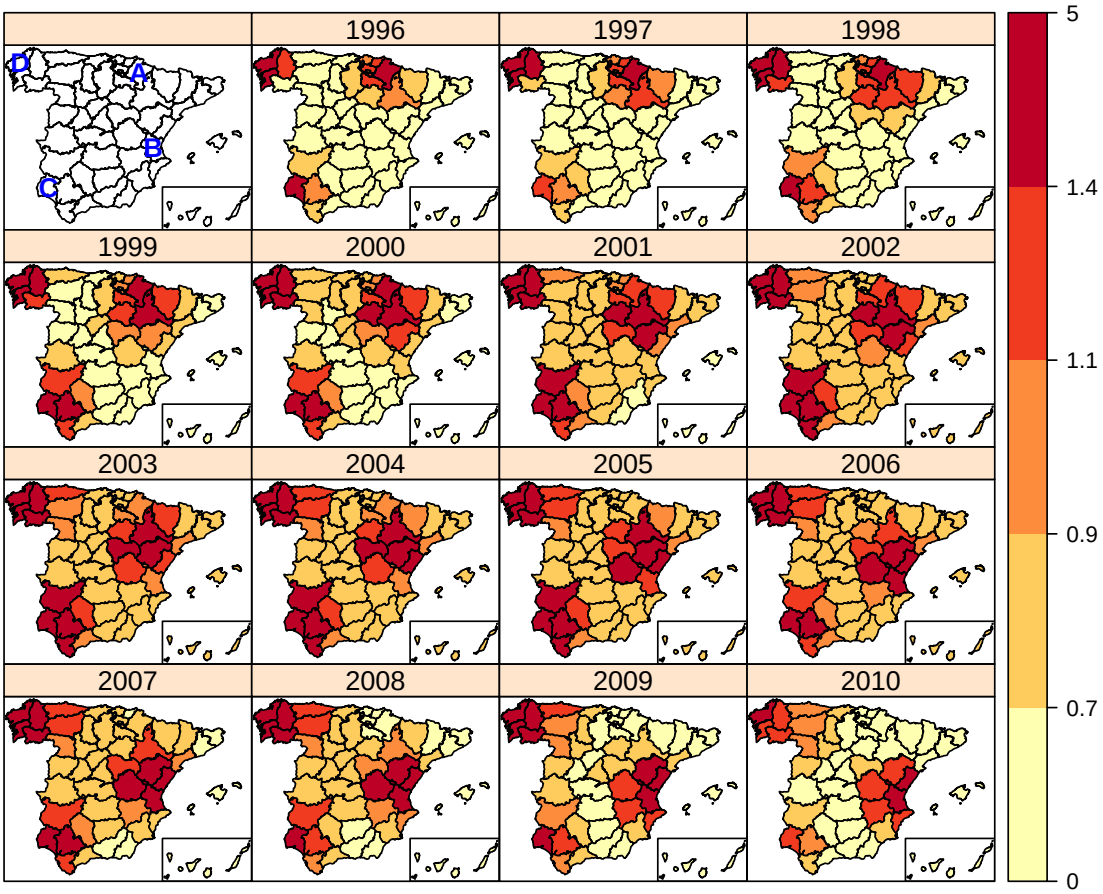


Figure 2.5: Scenario 2: true risk surface.

MRRMPSE are given in Table 2.4. We can se that the STMARS model behaves better in terms of MARB when estimating global risks. However, when the number of expected cases is very small ($E = 1$), the ANOVA-type P-spline model is almost as good as the STMARS model in terms of MARB. P-spline models have less MARB and MRRMPSE when estimating the temporal pattern. The MRRMPSE is also smaller when computing global risks and spatial patterns.

True and false positive rates are shown in Table 2.5. As in Scenario 1, P-spline models give, in general, high percentage of TPR. Again, although FPR are also slightly higher than in the rest of models, they are always below 6% even in the most extreme case ($E = 1$). In this model-free scenario, TPRs are higher and FPRs are smaller than in Scenario 1. This is sensible as the magnitude of the true high risks is larger in Scenario 2 than in Scenario 1, and consequently they can

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Table 2.4: Scenario 2: average values of MARB and MRRMPSE for relative risks and for spatial, temporal and spatio-temporal pattern estimates based on 100 simulated data sets.

	MARB				MRRMPSE			
	r_{it}	ξ_i^*	γ_t^*	δ_{it}^*	r_{it}	ξ_i^*	γ_t^*	δ_{it}^*
$E = 10$								
CAR model	0.0438	0.0132	0.0072	0.0410	0.1277	0.0781	0.0328	0.0918
Interaction P-spline	0.0605	0.0221	0.0050	0.0555	0.1130	0.0613	0.0254	0.0881
ANOVA P-spline	0.0536	0.0240	0.0071	0.0460	0.1071	0.0588	0.0229	0.0842
BYMar	0.0412	0.0123	0.0072	0.0402	0.1266	0.0770	0.0331	0.0922
STMARS	0.0395	0.0180	0.0073	0.0365	0.1216	0.0749	0.0333	0.0876
$E = 5$								
CAR model	0.0621	0.0227	0.0083	0.0556	0.1614	0.1049	0.0424	0.1080
Interaction P-spline	0.0847	0.0266	0.0049	0.0777	0.1449	0.0774	0.0331	0.1105
ANOVA P-spline	0.0702	0.0282	0.0056	0.0609	0.1373	0.0743	0.0290	0.1048
BYMar	0.0590	0.0172	0.0101	0.0561	0.1590	0.1020	0.0425	0.1095
STMARS	0.0565	0.0233	0.0106	0.0508	0.1523	0.0975	0.0432	0.1043
$E = 3$								
CAR model	0.0855	0.0384	0.0108	0.0657	0.1919	0.1284	0.0513	0.1189
Interaction P-spline	0.1020	0.0272	0.0094	0.0917	0.1721	0.0924	0.0400	0.1262
ANOVA P-spline	0.0824	0.0324	0.0075	0.0700	0.1637	0.0913	0.0369	0.1210
BYMar	0.0822	0.0325	0.0147	0.0691	0.1888	0.1239	0.0516	0.1222
STMARS	0.0765	0.0345	0.0155	0.0622	0.1800	0.1173	0.0515	0.1167
$E = 1$								
CAR model	0.1349	0.0812	0.0154	0.0897	0.2580	0.1863	0.0717	0.1417
Interaction P-spline	0.1428	0.0577	0.0076	0.1181	0.2360	0.1443	0.0541	0.1543
ANOVA P-spline	0.1276	0.0690	0.0105	0.0949	0.2343	0.1411	0.0536	0.1637
BYMar	0.1329	0.0689	0.0271	0.1031	0.2531	0.1745	0.0724	0.1542
STMARS	0.1240	0.0668	0.0259	0.0942	0.2407	0.1644	0.0703	0.1457

be detected more easily (Goicoa et al., 2012). Finally, as in Scenario 1, the ROC curves for the different models are plotted in Figure 2.6. The AUC values for the different models are almost identical in the scenario with the largest expected values ($E = 10$). However, slight differences are observed between models in the scenario with lowest expected cases ($E = 1$). Here the STMARS model, followed by the ANOVA-type P-spline model, outperforms the other models in terms of high-risk area detection.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Table 2.5: Scenario 2: true and false positives rates using lower one-sided credibility/confidence intervals.

	True positives			False positives		
	95% CI	90% CI	80% CI	95% CI	90% CI	80% CI
$E = 10$						
CAR model	0.631	0.703	0.787	0.001	0.004	0.012
Interaction P-spline	0.759	0.809	0.863	0.006	0.011	0.021
ANOVA P-spline	0.756	0.804	0.860	0.004	0.008	0.017
BYMar	0.613	0.684	0.772	0.001	0.003	0.010
STMARS	0.636	0.707	0.790	0.001	0.003	0.010
$E = 5$						
CAR model	0.531	0.612	0.715	0.002	0.005	0.015
Interaction P-spline	0.678	0.737	0.800	0.008	0.016	0.030
ANOVA P-spline	0.667	0.732	0.803	0.005	0.011	0.023
BYMar	0.506	0.589	0.692	0.001	0.004	0.012
STMARS	0.539	0.618	0.719	0.001	0.004	0.012
$E = 3$						
CAR model	0.439	0.530	0.646	0.002	0.005	0.017
Interaction P-spline	0.609	0.675	0.750	0.011	0.021	0.040
ANOVA P-spline	0.593	0.666	0.748	0.006	0.013	0.028
BYMar	0.412	0.500	0.620	0.001	0.004	0.013
STMARS	0.450	0.571	0.660	0.001	0.004	0.012
$E = 1$						
CAR model	0.244	0.333	0.467	0.001	0.004	0.018
Interaction P-spline	0.443	0.516	0.614	0.014	0.026	0.052
ANOVA P-spline	0.398	0.474	0.588	0.010	0.018	0.040
BYMar	0.215	0.297	0.425	0.001	0.003	0.012
STMARS	0.245	0.334	0.472	0.001	0.003	0.012

2.7 Discussion

In this chapter the performance of alternative space-time models in disease mapping has been compared varying the number of expected cases which, sometimes, is very small. This could be the case if rare diseases or very small domains like municipalities or census tracks are studied. In addition to analyzing bias and variability of the final risk estimates, both measures have been studied in the spatial, temporal and spatio-temporal effects separately. The variability of these separate patterns obtained

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

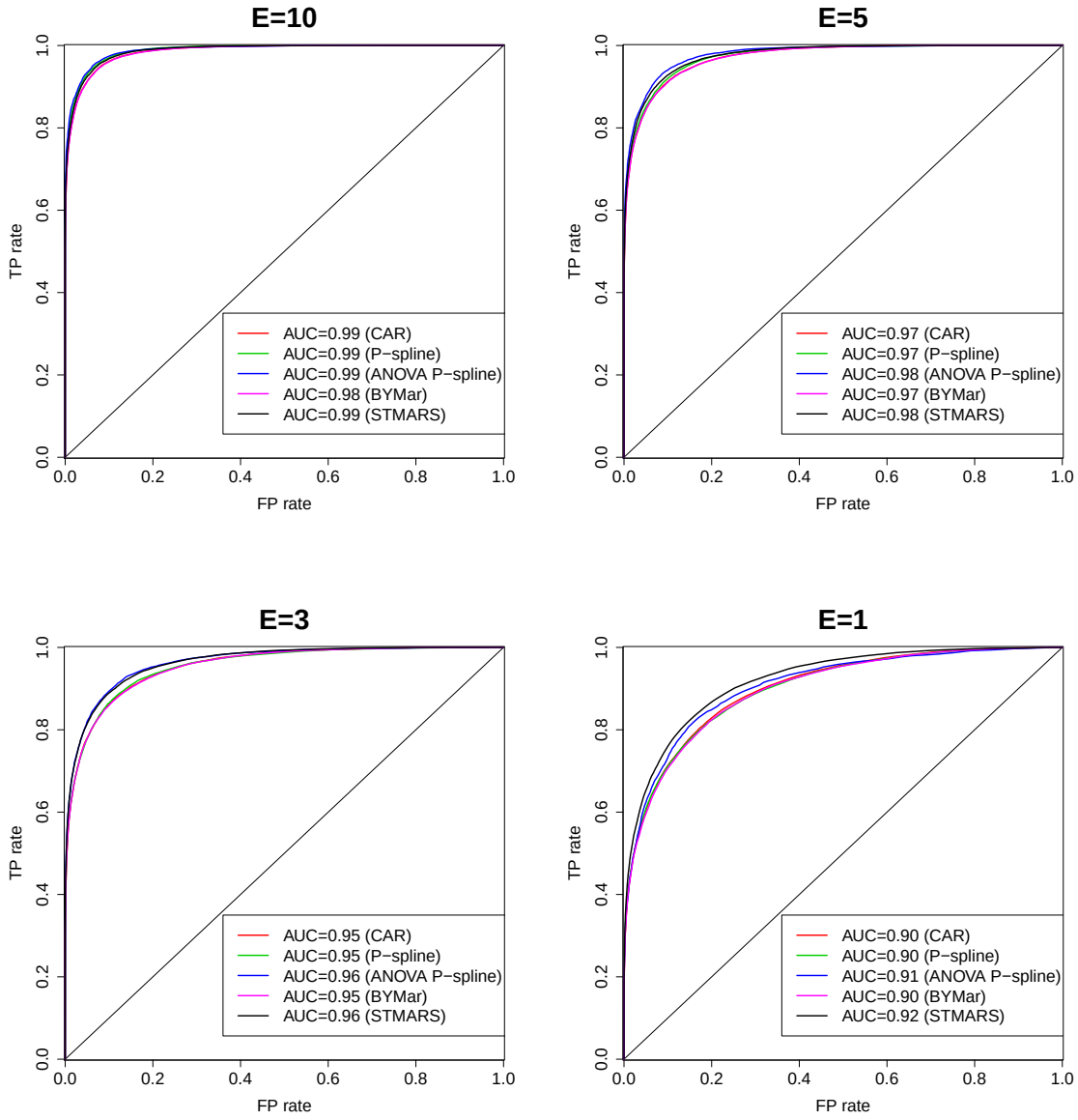


Figure 2.6: Scenario 2. AUC for different scale factor (SF) values.

from the P-spline models have been estimated using the delta method. Apart from using all the models using real data, two different simulation scenarios have been considered: one is based on the results obtained from a CAR model and the other one is a kind of "model-free" scenario. The true relative risks in Scenario 1 range from 0.63 to 1.65, whereas in Scenario 2 they vary between 0.51 and 3.71. This is a crucial difference as true high relative risks close to 1 are difficult to detect unless

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

the number of expected cases is big (Richardson et al., 2004; Goicoa et al., 2012; Ugarte et al., 2009b).

Results using the real data set seem to suggest that P-spline models are more rigid than random effect models. But against what at first sight might seem a possible disadvantage, this feature is sometimes an advantage i.e., when we have little information, the presumed "rigidity" of the P-spline models becomes an advantage because it allows a greater transfer of information between neighboring areas providing more stable risk estimates. The behavior of all the models improves in Scenario 2. This is expected because the magnitude of true high risks is greater, and detecting true positives depends on the risk value and on the number of expected cases, that is, the higher the risk is, the easier is to detect it. If the risk is moderately high, then a large number of expected cases is needed to detect it. Regarding false positives, in both scenarios they are under 10%, the value that can be considered acceptable. However, rates of false positives are smaller in Scenario 2. The reason is that in this scenario there are small true risks and they are seldom classified as true high risks. We note that a way to increase the rate of true positives is to reduce the confidence/credibility level (or equivalently to reduce the cutoff probability in Bayesian decision rules). This also increases the rates of false positives, but in any case it is greater than 10%. We have also computed the best rates of true positives obtained with the different models for fixed rates of false positives and we have observed that for the same rate of false positives, the confidence/credibility level (or equivalently the cutoff probability) needed to attain the best rate of true positives is different for each model. An interesting idea would be to use this information in the maps displaying the posterior probability that the relative risk is greater than one. Doing this, we can match colors to areas correctly classified with a high probability.

In summary, if the practitioner is particularly interested in finding high risk areas, we recommend the use of P-spline models as the rate of true positives is always higher than using the other models maintaining at the same time the percentage of false positives below 10%. P-spline models seem also to be better for estimating temporal trends. If the number of expected cases is very small, the ANOVA P-spline model could have some convergence problems due to the number of smoothing parameters. This fact could be alleviated using the same smoothing parameters for the interaction. In Chapter 4 a Bayesian implementation of three-dimensional P-spline models is proposed, avoiding convergence problems observed when fitting the models using the PQL technique.

The contents of this chapter have been published in Stochastic Environmental Research and Risk Assessment.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

3

Two-level spatially structured models in spatio-temporal disease mapping

3.1 Introduction

A new family of spatio-temporal models where the spatial effect has a two-level structure will be proposed in this chapter. These models will be useful for analyzing small area data that are grouped into larger regions. For example, we might be interested in studying mortality or incidence data in municipalities within provinces or health areas, or counties that are grouped into states. This grouping is fairly natural as sometimes the small areas within larger areas share the same health policies. As far as we know the first paper including two geographical-level random effects was published by MacNab and Dean (2000). Recently, multilevel regression models for geographical studies in sets of disjoint cities have been proposed (Marí-Dell'Olmo and Martínez-Beneito, 2015). However, all these models are restricted to the spatial case. The first paper we have found dealing with a two-level spatial random effect in spatio-temporal disease mapping is Schrödle et al. (2011). Two-level structure spatial effects are used to analyze reported cases of bovine viral diarrhoea in Switzerland because they suspected that the number of cases registered in each study area was highly influenced by the affiliation of those areas to a larger region structure. In consequence, besides the spatial random effect with an iCAR prior, an unstructured random effect was included to model the large region-level heterogeneity. Recently, similar models have been used by Ugarte et al. (2015a,b) to describe the temporal evolution of the geographical pattern of brain cancer incidence in Navarre and the Basque Country, and to analyze young people brain cancer mortality in Spanish provinces in the period 1986-2010, respectively. Our objective here is to propose some extensions of the previous models. These new model proposals include the four types of space-time interactions defined by Knorr-Held (2000) and allow for differ-

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

ent variance components in the covariance matrix of the spatial prior. Identifiability constraints are also derived for these new model proposals.

This chapter is structured in several sections. In Section 3.2 we introduce the new spatio-temporal models including two-level spatial random effects and the set of constraints are clearly established. Brain cancer mortality data in the municipalities of Navarre and the Basque Country are analyzed in Section 3.3, using the health areas as a second level of spatial aggregation. Finally, Section 3.4 assesses by simulation the performance of the new models in comparison with the non-parametric spatio-temporal models given by Knorr-Held (2000). The chapter concludes with a discussion and some conclusions.

3.2 Space-time models with two-level spatial random effects

The models described in the previous chapters are suitable for the analysis of spatio-temporal areal data in many real settings. However, these models may not be the best candidates if small areas are grouped into larger regions. For example, when analyzing mortality/incidence risks in municipalities that are aggregated in provinces with their own health systems, or when it is expected that small areas within the same health area share some common features (the same health policies for instance). In all theses cases, two-level spatial models are more appropriate. The two-level models also present advantages over the one-level models. They allow to identify regional effects at each level of spatial aggregation and to model spatio-temporal interactions at different levels. The former feature could help to model interpretation and the latter saves computational time when the interactions occur at the large area level, because the number of restrictions needed to identify the model can be greatly reduced. The new model proposals are explained in what follows.

Let us suppose now that the region under study is divided into n first-level areas (FLA) labeled as $i = 1, \ldots, n$ that can be aggregated into m second-level areas (SLA) labeled as $j = 1, \ldots, m$ where $m < n$. If data are available for several time periods $t = 1, \ldots, T$, the spatio-temporal model of Equation (1.4) can be extended using a SLA random effect (*two-level model A*) as

$$\log r_{it} = \eta + \xi_i + \psi_{j(i)} + \gamma_t + \delta_{it}, \tag{3.1}$$

where $j(i)$ denotes the SLA $j = 1, \ldots, m$ to which FLA i belongs. In this first extension we have FLA space-time interactions, but SLA space-time interactions

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

can be defined instead (*two-level model B*), i.e

$$\log r_{it} = \eta + \xi_i + \psi_{j(i)} + \gamma_t + \delta_{j(i)t}. \tag{3.2}$$

Note that SLA space-time interactions could be computationally more convenient. In general, a LCAR prior can be used for both spatial random effects, that is,

$$\begin{aligned} \boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)' &\sim N(\mathbf{0}, [\tau_\xi(\lambda_\xi \mathbf{R}_\xi + (1-\lambda_\xi)\mathbf{I}_n)]^{-1}), \\ \boldsymbol{\psi} = (\psi_1, \ldots, \psi_m)' &\sim N(\mathbf{0}, [\tau_\psi(\lambda_\psi \mathbf{R}_\psi + (1-\lambda_\psi)\mathbf{I}_m)]^{-1}), \end{aligned}$$

where $\mathbf{R}_\psi$ is the $m \times m$ spatial neighborhood matrix of the SLAs and $\mathbf{I}_m$ is an identity matrix of dimension $m \times m$. A first order random walk prior distribution has been considered here for the temporally structured random effect $\boldsymbol{\gamma}$. Finally, as for the usual one-level models, the different types of space-time interactions defined by Knorr-Held (2000) can be considered.

A sensible modification of Models A and B is to account for spatial variability only among those FLAs belonging to the same SLA. In this case, the FLA random effect is distributed as $\boldsymbol{\xi}^* \sim N(\mathbf{0}, [\tau_\xi \mathbf{D}_\xi^*]^-)$, and the structure matrix $\mathbf{D}_\xi^* = \lambda_\xi \mathbf{R}_\xi^* + (1-\lambda_\xi)\mathbf{I}_n$ has the following blocks

$$\mathbf{D}_\xi^* = \begin{pmatrix} \lambda_\xi \mathbf{R}_{\xi_1} + (1-\lambda_\xi)\mathbf{I}_{n_1} & & \\ & \ddots & \\ & & \lambda_\xi \mathbf{R}_{\xi_m} + (1-\lambda_\xi)\mathbf{I}_{n_m} \end{pmatrix}, \tag{3.3}$$

where $\mathbf{R}_\xi^* = \text{blockdiag}\,(\mathbf{R}_{\xi_1}, \ldots, \mathbf{R}_{\xi_m})$, $\mathbf{R}_{\xi_j}$ is the neighborhood matrix of FLAs within the jth SLA, $\mathbf{I}_{n_j}$ are the corresponding identity matrices, and n_j denotes the number of FLAs within the jth SLA. If we consider the structure matrix of Equation (3.3) and FLA interactions, the model will be named *two-level model C*. If SLA space-time interactions are considered instead, the model will be called *two-level model D*. These models are written as

$$\log r_{it} = \eta + \xi_i^* + \psi_{j(i)} + \gamma_t + \delta_{it}^*, \tag{3.4}$$

and

$$\log r_{it} = \eta + \xi_i^* + \psi_{j(i)} + \gamma_t + \delta_{j(i)t} \tag{3.5}$$

respectively. If completely structured (Type IV) interactions are used, the priors of the corresponding interaction terms of Equation (3.4) and Equation (3.5) are given by

$$\boldsymbol{\delta}^* \sim N(\mathbf{0}, [\tau_\delta(\mathbf{R}_\xi^* \otimes \mathbf{R}_\gamma)]^-) \qquad \text{and} \qquad \boldsymbol{\delta} \sim N(\mathbf{0}, [\tau_\delta(\mathbf{R}_\psi \otimes \mathbf{R}_\gamma)]^-).$$

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Table 3.1: Main characteristics of all models described in Section 3.2.

	FLA random effect	Space-time interaction (Type IV)	
		FLA: δ_{it}	SLA: $\delta_{j(i)t}$
Model A	$\boldsymbol{\xi} \sim N(\mathbf{0}, [\tau_\xi(\lambda_\xi \mathbf{R}_\xi + (1-\lambda_\xi)\mathbf{I}_n)]^{-1})$	$\boldsymbol{\delta} \sim N(\mathbf{0}, [\tau_\delta(\mathbf{R}_\xi \otimes \mathbf{R}_\gamma)]^-)$	–
Model B		–	$\boldsymbol{\delta} \sim N(\mathbf{0}, [\tau_\delta(\mathbf{R}_\psi \otimes \mathbf{R}_\gamma)]^-)$
Model C	$\boldsymbol{\xi}^* \sim N(\mathbf{0}, [\tau_\xi(\lambda_\xi \mathbf{R}^*_\xi + (1-\lambda_\xi)\mathbf{I}_n)]^{-1})$	$\boldsymbol{\delta}^* \sim N(\mathbf{0}, [\tau_\delta(\mathbf{R}^*_\xi \otimes \mathbf{R}_\gamma)]^-)$	–
Model D		–	$\boldsymbol{\delta} \sim N(\mathbf{0}, [\tau_\delta(\mathbf{R}_\psi \otimes \mathbf{R}_\gamma)]^-)$
Model E	$\boldsymbol{\xi}_j \sim N(\mathbf{0}, [\tau_{\xi_j}(\lambda_{\xi_j} \mathbf{R}_{\xi_j} + (1-\lambda_{\xi_j})\mathbf{I}_{n_j})]^{-1})$	$\boldsymbol{\delta}^* \sim N(\mathbf{0}, [\tau_\delta(\mathbf{R}^*_\xi \otimes \mathbf{R}_\gamma)]^-)$	–
Model F	for $j = 1, \ldots, m$	–	$\boldsymbol{\delta} \sim N(\mathbf{0}, [\tau_\delta(\mathbf{R}_\psi \otimes \mathbf{R}_\gamma)]^-)$

FLA: first-level area; SLA: second level area.

Further extensions of Models C and D (named *two-level model E* -for FLA interactions- and *two-level model F* -for SLA interactions-) are proposed defining m independent random effects to model the spatial dependence within each SLA, so that different variance components are attached to the blocks of the inverse of the structure matrix $\mathbf{D}^*_\xi$. That is, decomposing the spatial random effect as $\boldsymbol{\xi}^* = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m)'$, we assume that

$$\boldsymbol{\xi}_j = (\xi_{j1}, \ldots, \xi_{jn_j})' \sim N\left(\mathbf{0}, [\tau_{\xi_j}(\lambda_{\xi_j} \mathbf{R}_{\xi_j} + (1-\lambda_{\xi_j})\mathbf{I}_{n_j})]^{-1}\right), \text{ for } j = 1, \ldots, m.$$

The main characteristics of all models described in this section are summarized in Table 3.1 for completely structured interactions. The specification of the other types of interactions is similar, just substituting the structure matrix by the corresponding identity matrix if there is no structure in space or time.

Two-level models A and B can be seen as extensions of the one proposed by Schrödle et al. (2011) because these authors propose independent SLA effects. Two-level models C and D are spatio-temporal extensions of the models proposed by MacNab and Dean (2000), and models E and F consider similar spatial structures as both Models 1 and 2 in Marí-Dell'Olmo and Martínez-Beneito (2015), and the models used in Ugarte et al. (2015a,b).

Similar to the spatio-temporal CAR models described in Chapter 1, sum-to-zero constraints are needed to avoid identifiability problems between the intercept, the main effects, and the interaction term. Table 3.2 summarizes the necessary constraints if a LCAR (or iCAR) prior distribution is considered for both FLA and SLA random effects ξ_i and $\psi_{j(i)}$, a RW1 prior distribution is considered for the temporal random effect γ_t, and the spatio-temporal random effects δ_{it} (or $\delta_{j(i)t}$) are completely structured (Type IV interaction). Full details about how these constraints are derived are provided in Appendix 3A.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Table 3.2: Identifiability constraints for the space-time models including two-level spatial random effects described in Section 3.2 and Type IV space-time interactions.

	FLA interaction	SLA interaction
Model A/B	$\sum_{i=1}^{n}\xi_i=0,\quad \sum_{j=1}^{m}\psi_j=0,\quad \sum_{t=1}^{T}\gamma_t=0$ $\sum_{t=1}^{T}\delta_{it}=0,\quad \text{for}\quad i=1,\ldots,n$ $\sum_{i=1}^{n}\delta_{it}=0,\quad \text{for}\quad t=1,\ldots,T$	$\sum_{i=1}^{n}\xi_i=0,\quad \sum_{j=1}^{m}\psi_j=0,\quad \sum_{t=1}^{T}\gamma_t=0$ $\sum_{t=1}^{T}\delta_{jt}=0,\quad \text{for}\quad j=1,\ldots,m$ $\sum_{j=1}^{m}\delta_{jt}=0,\quad \text{for}\quad t=1,\ldots,T$
Model C/D	$\sum_{i=1}^{n_j}\xi_{j,i}=0,\quad \text{for}\quad j=1,\ldots,m$ $\sum_{j=1}^{m}\psi_j=0,\quad \sum_{t=1}^{T}\gamma_t=0,$ $\sum_{t=1}^{T}\delta_{it}=0,\quad \text{for}\quad i=1,\ldots,n$ $\sum_{i=1}^{n_j}\delta_{j,it}=0,\quad \text{for}\quad j=1,\ldots,m;\ t=1,\ldots,T$	$\sum_{i=1}^{n_j}\xi_{j,i}=0,\quad \text{for}\quad j=1,\ldots,m$ $\sum_{j=1}^{m}\psi_j=0,\quad \sum_{t=1}^{T}\gamma_t=0,$ $\sum_{t=1}^{T}\delta_{jt}=0,\quad \text{for}\quad j=1,\ldots,m$ $\sum_{j=1}^{m}\delta_{jt}=0,\quad \text{for}\quad t=1,\ldots,T$
Model E/F	Same constraints as models C and D respectively	

FLA: first-level area; SLA: second level area.

3.3 Illustration

In this section brain cancer mortality data in the municipalities of Navarre and the Basque Country (Spain) will be analyzed to illustrate the new model proposals. Health areas will be considered as the second level of spatial aggregation in order to identify potential different temporal evolutions for each health area. The `R` code to fit some of these models using the `R-INLA` library is also described in Appendix 3B.

3.3.1 Brain cancer mortality data in the municipalities of Navarre and Basque Country

The study covers all brain cancer deaths (International Classification of Diseases-10: code C71) registered in Navarre and Basque Country during the period 1986-2008.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

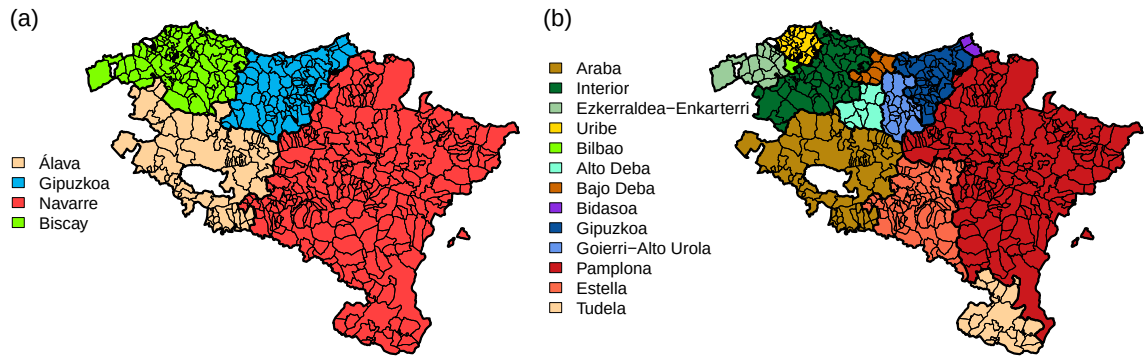


Figure 3.1: Map of the $n = 501$ municipalities of Navarre and Basque Country. (a) Municipalities grouped in provinces; (b) Municipalities aggregated by SLAs.

There is a total of 501 municipalities in which 3598 cases were recorded throughout the study period (353 in Álava, 929 in Gipuzkoa, 922 in Navarre, and 1394 in Biscay). Age and sex-standardization has been used to compute the number of expected cases (internal standardization). This method allows us to compare each municipality in a certain year with the overall area throughout the entire study period. The expected cases range from 0 to 25, while the number of observed cases varies from 0 to 35.

The two-level models described in Section 3.2 have been fitted accounting for the $n = 501$ municipalities as FLAs and the $m = 13$ health areas of Navarre and the Basque Country as SLAs (see Figure 3.1). Models including only municipality effects have been also studied to investigate whether or not a health area effect is necessary in the model.

To fit the models a non-informative uniform prior distribution on the positive real line was given to all the standard deviations. A Uniform(0,1) distribution for the smoothing parameters of the LCAR priors have been also considered. Results from fitting all the models in INLA (using the Gaussian approximation strategy) are displayed in Table 3.3. As the models with FLA interactions were not better than those with SLA space-time interactions, two-level model E has not been fitted in this analysis.

The DIC, the corrected DIC, and the logarithmic score suggest a two-level model B with a Type II interaction as the best model. We recall that this model has a municipal-level spatial random effect and a health area-level spatial effect, both of them with a LCAR prior distribution. A random walk of order one (RW1) is considered as a prior for the temporal effect, and spatially unstructured temporal effects (Type II) are selected for the SLA space-time interaction term. The model

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Table 3.3: Model comparisons in the analysis of brain cancer mortality data in Navarre and the Basque Country.

	$\gamma \equiv$ Random walk of first order						
	Space-time interaction	$\overline{D}$	p_D	DIC	DICc	LS	time
CAR model	Type I	8575.2	101.9	8677.1	8687.1	4339.0	1
	Type II	8594.4	80.7	8675.3	8679.2	4338.0	47
	Type III	8582.3	92.9	8675.2	8681.3	4338.0	4
	Type IV	8600.6	71.8	8672.3	8675.1	4336.6	90
Two-level model A	Type I	8583.5	97.7	8681.1	8690.7	4341.0	<1
	Type II	8602.3	77.1	8679.5	8683.2	4340.2	50
	Type III	8591.1	88.1	8679.1	8684.9	4340.0	2
	Type IV	8608.9	67.6	8676.5	8679.0	4338.7	65
Two-level model B	Type I	8589.7	86.1	8675.8	8683.3	4338.4	<1
	Type II	**8603.3**	**67.3**	**8670.6**	**8673.9**	**4336.1**	<1
	Type III	8599.2	77.8	8677.0	8682.7	4339.2	<1
	Type IV	8611.6	62.6	8674.3	8676.8	4337.7	<1
Two-level model C	Type I	8588.5	96.3	8684.8	8694.0	4342.9	1
	Type II	8606.3	77.0	8683.2	8687.1	4342.1	54
	Type III	8605.3	82.3	8687.6	8691.3	4343.7	14
	Type IV	8621.0	64.8	8685.8	8687.9	4343.1	161
Two-level model D	Type I	8592.6	86.8	8679.5	8687.2	4340.3	<1
	Type II	8607.0	68.4	8675.5	8678.9	4338.1	<1
	Type III	8603.0	77.7	8680.7	8686.5	4341.1	<1
	Type IV	8615.1	62.8	8678.0	8680.6	4339.6	<1
Two-level model F	Type I	8650.2	30.3	8680.5	8681.0	4340.9	14
	Type II	8650.9	28.8	8679.7	8680.3	4340.4	14
	Type III	8649.0	31.3	8680.3	8680.9	4340.8	15
	Type IV	8648.3	31.2	8679.5	8680.0	4340.4	16

DIC: deviance information criterion, DICc: corrected version of deviance information criterion, $\overline{D}$: mean deviance, p_D: effective number of parameters, LS: logarithmic score, and time: computational time in minutes.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Table 3.4: Estimated posterior means and 95% credibility intervals for model parameters using full Laplace approximation.

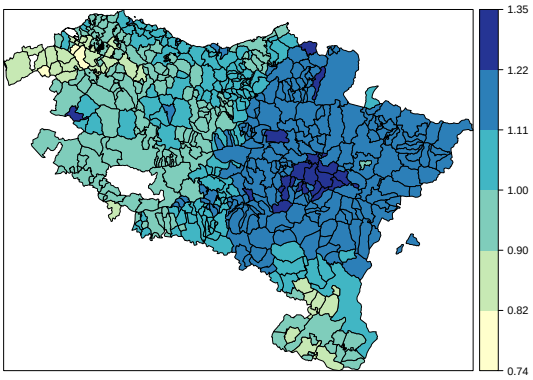
Two-level model B (Type II) $\log r_{it} = \eta + \xi_i + \psi_{j(i)} + \gamma_t + \delta_{j(i)t}$			
	Mean	2.5 %	97.5 %
η	-0.023	-0.081	0.036
τ_ξ	24.6	6.7	69.6
λ_ξ	0.589	0.126	0.947
τ_ψ	92.3	12.1	369.3
λ_ψ	0.466	0.066	0.904
τ_γ	358.3	70.5	1131.7
τ_δ	498.6	130.2	1531.7

has been fitted again using the full Laplace approximation strategy.

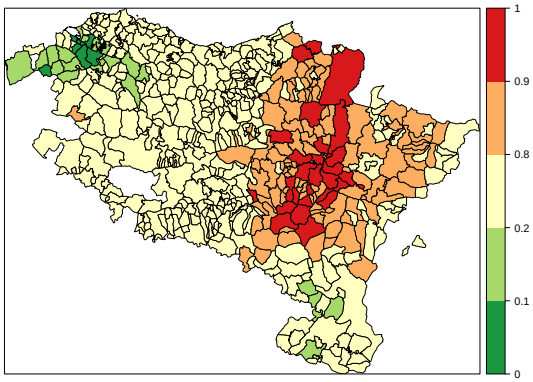
The percentage of variability of the overall risk explained by the health area-level interaction component $\hat{\delta}_{j(i)t}$ is about 8%, while most of the variability is explained by the temporal effect $\hat{\gamma}_t$ (about 62% of the total variability). Estimated posterior means and 95% credibility intervals for the model parameters are given in Table 3.4. We notice that the spatial smoothing at municipality level seems to be stronger than that at health area level.

The estimated spatial and temporal patterns are shown in Figure 3.2. The spatial mortality risk pattern $\exp(\hat{\xi}_i + \hat{\psi}_{j(i)})$ associated to each municipality (constant during the whole studied period), as well as posterior probabilities that these risks are greater than one are displayed in Figure 3.2a and Figure 3.2b, respectively. The temporal pattern common to all regions is also visualized on Figure 3.2c, where a global increase is observed. Space-time interactions for SLAs are given in Figure 3.3. There, one can observe how the health area specific temporal trends differ. While some of them like Uribe and Ezkerraldea-Enkarterri contribute to increase the final risk, others like Gipuzkoa or Goierri-Alto Urola do the opposite. Finally, the temporal evolution of the estimated brain cancer mortality risks for the most populated municipalities of each health area in Navarre and the Basque Country are given in Figure 3.4. The red color indicates a high posterior probability of relative risks being greater than one.

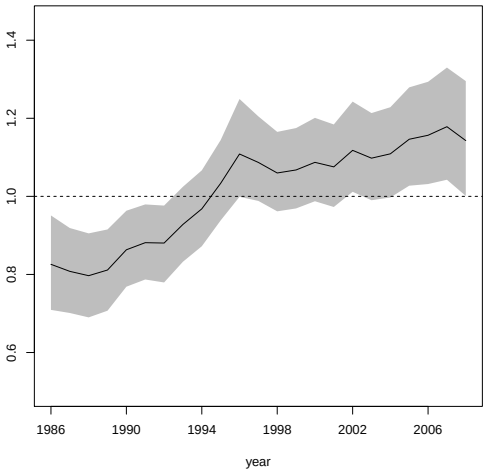
upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa



(a) Map of the spatial pattern of mortality risks $\exp(\hat{\xi}_i + \hat{\psi}_{j(i)})$



(b) Map of posterior probabilities of the spatial pattern of mortality risk $P(\exp(\hat{\xi}_i + \hat{\psi}_{j(i)}) > 1|\mathbf{O})$



(c) Temporal trend of brain cancer mortality relative risk $\exp(\hat{\gamma}_t)$

Figure 3.2: Global spatial and temporal patterns in the analysis of brain cancer mortality in Navarre and the Basque Country.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

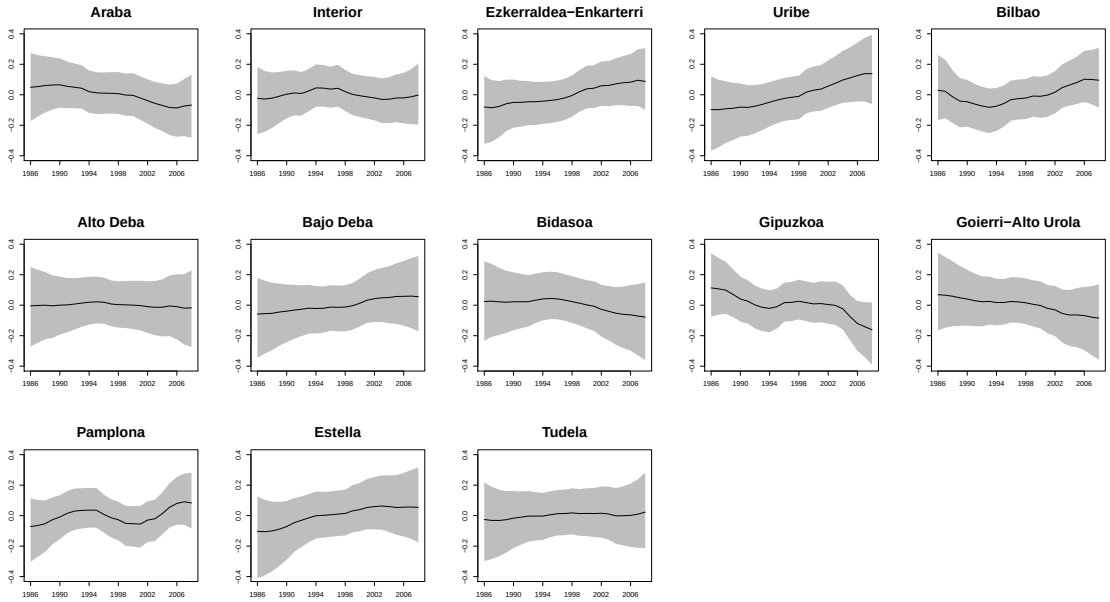


Figure 3.3: Space-time interaction term $\hat{\delta}_{j(i)t}$ for each health area.

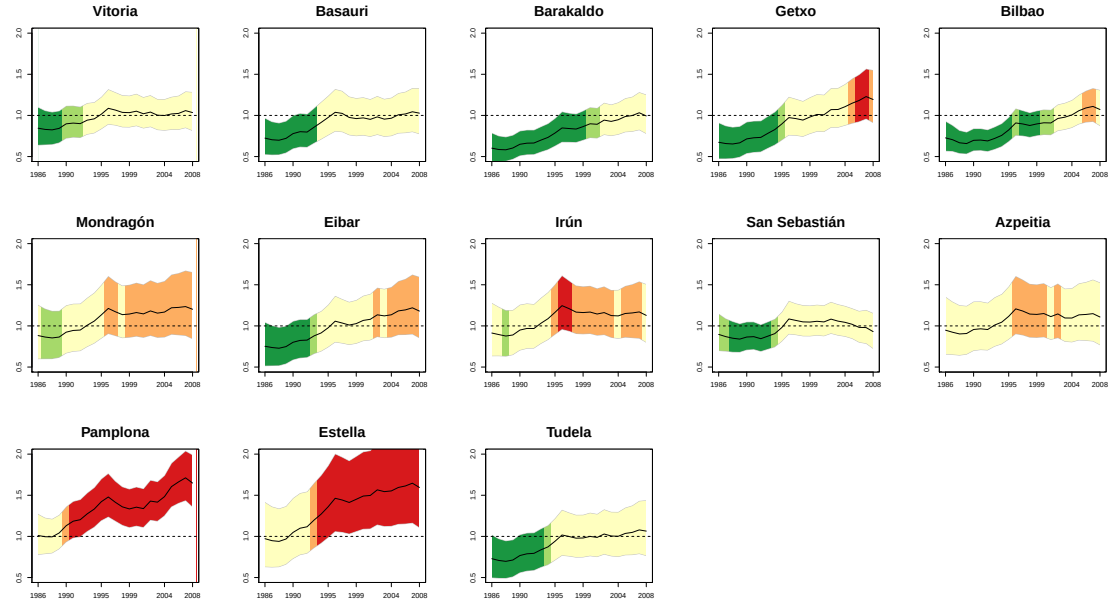


Figure 3.4: Temporal evolution of the estimated brain cancer mortality risks $\hat{r}_{it}$ for the more populated municipalities of each heath area in Navarre and the Basque Country (Spain) and 95 % two-sided credible intervals. The red color indicates a high posterior probability of relative risks being greater than one.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

3.4 Simulation study

In this section, a simulation study is conducted to compare the performance of the models introduced in Section 3.2 and the commonly used spatio-temporal CAR models (see Section 1.2). We base our study on the $n = 501$ municipalities of the Autonomous Regions of Navarre and the Basque Country, Spain. To imitate the real case study that is analyzed in the previous section, the number of time periods (years) is $T = 20$. Two different scenarios have been created. In Scenario 1, the municipalities are grouped in $m = 4$ provinces called Álava, Gipuzkoa, Navarre, and Biscay. To evaluate the effect of increasing the number of SLAs, municipalities are aggregated into $m = 13$ health areas in Scenario 2 (see Figure 3.1).

3.4.1 Data generation

Counts are generated using a Poisson distribution with mean $e_{it}r_{it}$ where the expected number of cases is 1, 3, 5 or 10 giving rise to four different subscenarios. A common true spatio-temporal log-risk surface has been generated as the sum of a FLA random term (ξ_i), a SLA geographical effect ($\psi_{j(i)}$) -provinces in Scenario 1 and health areas in Scenario 2-, a global temporal effect (γ_t) and a SLA space-time interaction effect ($\delta_{j(i)t}$) i.e.,

$$\log r_{it} = \xi_i + \psi_{j(i)} + \gamma_t + \delta_{j(i)t} \quad \text{for } j = 1, \ldots, m; \quad t = 1, \ldots, T. \tag{3.6}$$

Scenario1

To create a scenario where the geographical distribution of risks varies among the provinces, $m = 4$ independent random effects have been generated from a LCAR distribution with different amount of spatial variability, i.e, the spatial random effect $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\xi}_3, \boldsymbol{\xi}_4)'$ has been generated as

$$\boldsymbol{\xi}_j \sim N(\mathbf{0}, [\tau_{\xi_j}(\lambda_{\xi_j}\mathbf{R}_{\xi_j} + (1 - \lambda_{\xi_j})\mathbf{I}_{n_j})]^{-1}), \quad \text{for} \quad j = 1, \ldots, 4$$

where $\mathbf{R}_{\xi_j}$ is the spatial neighborhood matrix of municipalities within each province, $\mathbf{I}_{n_j}$ is the identity matrix of size n_j (number of municipalities in province j) and λ_{ξ_j} takes the values $\{0.1, 0.4, 0.6, 0.9\}$. The precision components τ_{ξ_j} have been chosen so that the random effects $\boldsymbol{\xi}_j$ take values in the range $[-0.415, 0.405]$, which is equivalent to $\exp(\boldsymbol{\xi}_j) \in [0.66, 1.5]$. Note that 1.5 means that the risk of that area is 50% higher than in the overall study area, and 0.66 (1/1.5) means the same 50% but lower than the overall study area.

As only four SLA are considered in this scenario, the province level random effect $\psi_{j(i)}$ has been fixed in order to keep the simulated log-risks into a controlled range.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

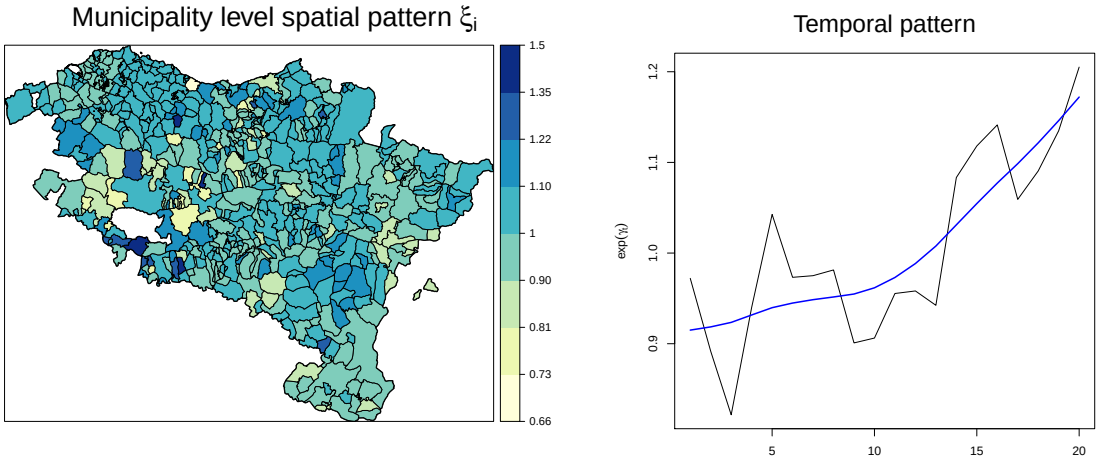


Figure 3.5: Map of the municipality level spatial patterns of mortality risk (left) and global temporal pattern (right) in Scenario 1.

Specifically, it has been set up as

$$\boldsymbol{\psi} = (0.2 \times \mathbf{1}'_{n_1}, -0.2 \times \mathbf{1}'_{n_2}, 0.1 \times \mathbf{1}'_{n_3}, -0.1 \times \mathbf{1}'_{n_4})',$$

where $\mathbf{1}_{n_j}$ are vector of ones of length n_j, i.e., the random effect is defined so that municipalities within the same province take the same value.

The temporal random effect γ_t has been randomly generated from a first order random walk

$$\boldsymbol{\gamma} \sim N(\mathbf{0}, [\tau_\gamma \mathbf{R}_\gamma]^-),$$

where $\mathbf{R}_\gamma$ is the $T \times T$ structure matrix of a RW1. The precision component τ_γ has been chosen so that γ_t takes values in the range $[-0.415, 0.405]$, similar to the spatial case. To imitate a temporal pattern corresponding to a possible real scenario, a smooth version of this random effect has been considered. Both spatial and temporal patterns are shown in Figure 3.5. Finally, province level space-time interactions $\delta_{j(i)t}$ have been generated from different parametric trend shapes for each province (see Figure 3.6).

Scenario2

In this scenario, the number of SLAs has been increased to $m = 13$, corresponding to the number of health areas of Navarre and the Basque Country. Similarly to Scenario 1, m independent LCAR distribution random effects have been generated

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

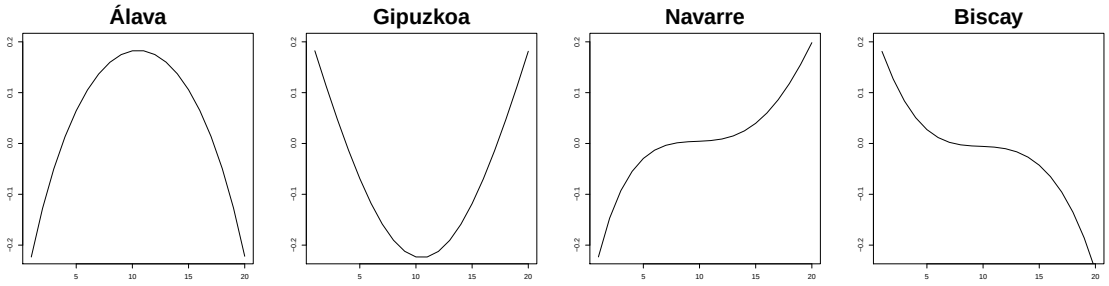


Figure 3.6: Province level space-time interactions $\delta_{j(i)t}$ for the generated log-risks surface in Scenario 1.

for the municipality level spatial random effect $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_m)'$, where λ_{ξ_j} takes the same values as in Scenario 1, $\{0.1, 0.2, 0.6, 0.9\}$, in such a way that health areas within the same province have a common value. However, unlike Scenario 1, a LCAR distribution with a spatial smoothing parameter equal to $\lambda_\psi = 0.75$ has been considered for the SLA random effect. Once again, the precision components τ_{ξ_j} and τ_ψ have been chosen so that random effects $\boldsymbol{\xi}_j$ and $\boldsymbol{\psi}$ take values in the range [-0.415,0.405]. The sum of both spatial patterns is plotted on the left in Figure 3.7. For the global temporal pattern, the same random effect γ_t defined for Scenario 1 has been used. Finally, specific temporal trends have been generated for each health area to determine the space-time interaction term $\delta_{j(i)t}$ (see right hand side plot in Figure 3.7). These second-order polynomial trends change gradually from a U-shaped curve for the north-westernmost region (Ezkerraldea-Enkarterri) to an inverse-U shaped curve for the south-easternmost health area (Tudela).

The final true spatio-temporal risk surfaces defined in Equation (3.6) for both Scenarios 1 and 2 are shown in Figure 3.8. The amount of variability explained by the interaction term is about 20% in both cases. The simulated counts are generated from a Poisson distribution with mean $\mu_{it} = e_{it} r_{it}$ giving rise to eight different subscenarios (four different expected values for each scenario). For each subscenario thirty data sets have been generated and five different models have been fitted: a spatio-temporal LCAR model (see Equation (1.4)) and four two-level models; in particular, a two-level model A (with FLA space-time interactions), two-level models B and D (with SLA interactions), and finally, a two-level structure model F with m independent spatial random effects, $\boldsymbol{\xi}_j \sim N(\mathbf{0}, [\tau_{\xi_j}(\lambda_{\xi_j}\mathbf{R}_{\xi_j} + (1-\lambda_{\xi_j})\mathbf{I}_{n_j})]^{-1})$ (with SLA interactions). Models C and E are not considered because after fitting models A and B, model B (with SLA interactions) is always better (see Table 3.5). In all these models, a LCAR and a RW1 prior have been used for the municipality-level spatial random effect ξ_i and the structured temporal effect γ_t, respectively. Type II and

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

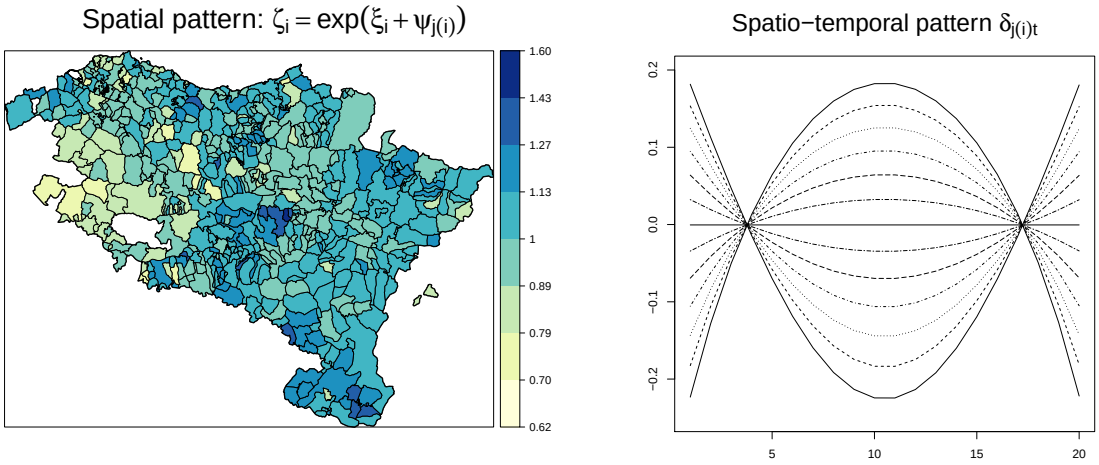


Figure 3.7: Map of the sum of municipality level ξ_i and basic health area level $\psi_{j(i)}$ spatial patterns of mortality risk (left) and SLA space-time interactions $\delta_{j(i)t}$ (right) for the generated log-risks surface in Scenario 2.

Type IV space-time interactions have been fitted for both municipality and province level interactions. In Type II interactions, temporal trends are different from region to region but do not have any structure in space while in Type IV interactions the temporal trends are likely to be similar for adjacent regions. Finally, for SLA spatial random effects $\psi_{j(i)}$, an exchangeable prior $N(\mathbf{0}, \tau_\psi^{-1}\mathbf{I}_m)$ has been considered for the province level random effect in Scenario 1, while a LCAR prior has been used for the health area level random effect in Scenario 2.

3.4.2 Results

Several measures have been computed to evaluate the performance of the two-level models proposed in Section 3.2. Apart from the Deviance Information Criterion (DIC), the corrected DIC version (DICc) described in Equation (1.13) has also been calculated. In addition, conditional predictive ordinates (CPO) have been computed as they facilitate the computation of the cross-validated logarithmic score (LS) described in Equation (1.14). Finally, root mean squared errors (RMSE) and mean absolute bias (MAB) have also been computed. Namely,

$$RMSE = \sqrt{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(\hat{r}_{it} - r_{it})^2} \quad \text{and} \quad MAB = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}|\hat{r}_{it} - r_{it}|.$$

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

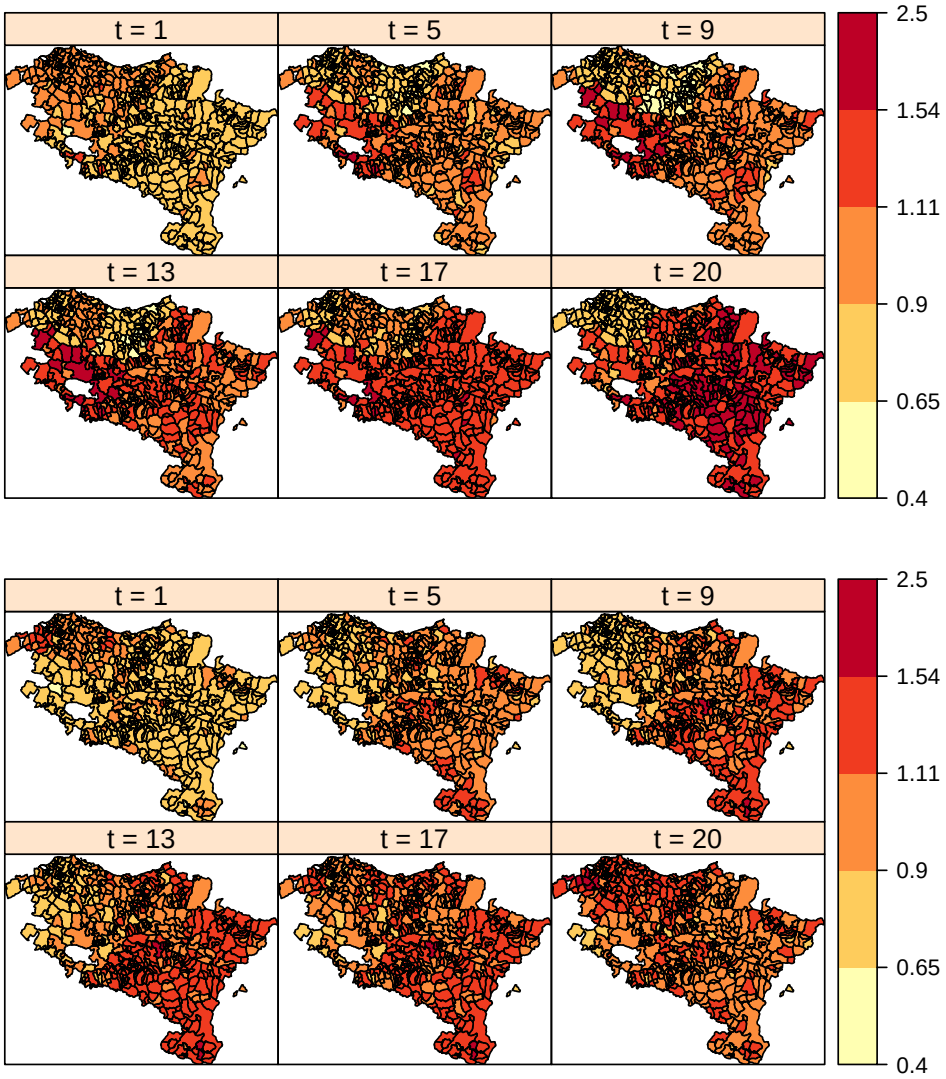


Figure 3.8: Spatio-temporal true risk surface for the simulation study of Scenario 1 (top panel) and Scenario 2 (bottom panel).

For all these criteria, lower values imply better model properties. The average values for the thirty simulated data sets in each of the subscenarios have been computed in Table 3.5. In addition, average coverage percentages of the 95% credible intervals for the risks as well as true (TPR) and false positive rates (FPR) have been calculated. TPR (expressed in percentages) have been computed as the proportion of high true risks ($r_{it} > 1$) that were classified as high risks (a risk is classified as a high risk if the lower bound of the corresponding upper one-sided 95% credible

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

interval is greater than one). FPR were similarly computed. In Scenario 1, models with Type II interactions provide the best fitting, while Type IV interactions are the most suitable in Scenario 2. This is reasonable taking into account how the spatio-temporal patterns are defined to generate the log-risk surfaces. To conserve space, Table 3.5 only shows results obtained with these interactions.

We observe that the performance of the two-levels models is generally better than the single level spatial model. In addition, the single-level model and two-level model A are computationally expensive compared to models with SLA interactions since the amount of identifiability constraints needed in the interaction term $\boldsymbol{\delta}$ depends on the number of study regions ($n = 501$ in municipality level interaction models and $m = 4$ or $m = 13$ in province or health area level interaction models, respectively). When the number of expected cases increases, the model selection criteria discriminate better among all the models.

In Scenario 1, two-level Model F is generally selected as the best candidate for all possible values of expected cases. Although the average coverage probability is low when the number of expected cases is small, the sensitivity (TPR) and 1-specificity (FPR) of Model F are better than those of the single-level model. In Scenario 2, Model D is the best candidate in the majority of the cases, except when the number of expected cases is 10, where Model F seems to be slightly better. If the number of expected cases is small and the number of SLA grows, it is difficult to select Model F as the best candidate even when simulating under this model. It seems that there is not enough data to estimate this rather complex model. Although the performance of Model F in terms of average coverage probabilities is bad when the number of expected cases is very small, the model is the best in terms of TPR. If the number of expected cases increases, the behavior of this model improves greatly. The apparent good behavior of the single-level model in Scenario 2 when $E = 1$ in terms of average coverage probabilities is caused by the average length of the credible interval which almost doubles the average length of the credible interval in Model F.

This simulation study shows that the two-level spatially structured random effect models are better candidates than the single-level models when analyzing small area data grouped in larger areas like provinces or health areas. Health policies affecting the small areas within SLAs are often very similar and this is the main reason why it seems sensible to use two-level spatial models.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Table 3.5: Average values of DIC, corrected DIC, logarithmic score (LS), root mean squared errors (RMSE), mean absolute bias (MAB), average coverage percentages of the 95% credible interval for the risks, true positive rates (TPR) and false positive rates (FPR) in Scenario 1 and Scenario 2.

		Scenario 1 ($m = 4$)				**Scenario 2** ($m = 13$)			
		Type II interaction				Type IV interaction			
		E=1	E=3	E=5	E=10	E=1	E=3	E=5	E=10
DIC	CAR model	26519	39289	44969	52493	26546	39187	44758	52024
	TL model A	26479	39227	44901	52431	26539	39173	44763	52010
	TL model B	26392	38973	44514	51808	26529	39149	44704	51945
	TL model D	26400	38980	44521	51794	26523	39146	44696	51932
	TL model F	26389	38947	44476	51740	26536	39156	44700	51925
DICc	CAR model	26522	39311	45015	52614	26549	39198	44778	52062
	TL model A	26480	39243	44940	52541	26542	39184	44763	52047
	TL model B	26393	38978	44522	51808	26530	39155	44713	51961
	TL model D	26400	38983	44528	51806	26524	39150	44703	51946
	TL model F	26390	38951	44482	51751	26537	39159	44706	51938
LS	CAR model	13260	19650	22496	26280	13273	19594	22381	26016
	TL model A	13240	19628	22461	26247	13269	19587	22373	26009
	TL model B	13196	19487	22258	25900	13264	19576	22354	25976
	TL model D	13200	19490	22262	25900	13262	19574	22350	25969
	TL model F	13195	19474	22239	25873	13268	19579	22352	25966
RMSE	CAR model	0.159	0.135	0.123	0.105	0.114	0.095	0.083	0.068
	TL model A	0.145	0.128	0.118	0.102	0.112	0.092	0.081	0.066
	TL model B	0.109	0.087	0.076	0.061	0.108	0.088	0.076	0.062
	TL model D	0.114	0.089	0.077	0.062	0.107	0.086	0.075	0.061
	TL model F	0.109	0.081	0.070	0.057	0.114	0.093	0.079	0.061
MAB	CAR model	0.120	0.102	0.093	0.080	0.087	0.073	0.064	0.053
	TL model A	0.107	0.095	0.088	0.077	0.086	0.071	0.063	0.052
	TL model B	0.079	0.065	0.057	0.047	0.083	0.068	0.059	0.048
	TL model D	0.081	0.066	0.058	0.047	0.082	0.066	0.058	0.048
	TL model F	0.079	0.061	0.053	0.044	0.087	0.071	0.060	0.048
Coverage (%)	CAR model	88.3	91.9	92.8	94.4	95.8	95.8	96.3	96.8
	TL model A	84.8	92.0	93.3	94.5	94.5	96.6	97.0	97.2
	TL model B	94.2	95.3	95.5	95.5	91.1	94.8	95.2	95.6
	TL model D	87.0	93.1	94.0	94.8	87.9	92.8	94.0	94.9
	TL model F	84.8	93.3	94.0	93.8	73.6	84.1	90.7	94.2
TPR (sensitivity)	CAR model	0.346	0.412	0.467	0.525	0.147	0.300	0.384	0.491
	TL model A	0.492	0.472	0.505	0.546	0.203	0.299	0.382	0.490
	TL model B	0.514	0.583	0.631	0.695	0.305	0.396	0.477	0.575
	TL model D	0.593	0.616	0.652	0.706	0.403	0.469	0.526	0.600
	TL model F	0.567	0.603	0.642	0.707	0.522	0.538	0.550	0.617
FPR (1-specificity)	CAR model	0.014	0.008	0.006	0.005	0.002	0.004	0.003	0.003
	TL model A	0.017	0.008	0.007	0.005	0.067	0.004	0.003	0.003
	TL model B	0.009	0.007	0.006	0.005	0.015	0.007	0.006	0.005
	TL model D	0.018	0.010	0.008	0.006	0.023	0.011	0.008	0.006
	TL model F	0.015	0.006	0.004	0.004	0.059	0.030	0.015	0.008

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

3.5 Discussion and conclusions

The statistical models used in space-time disease mapping have, in general, a single level of spatial dependence which means that a single-level area random effect is used. However, in some cases it makes sense to consider natural groupings of the small areas. For example, municipalities can be grouped into provinces or health areas, or counties can be aggregated in states affected by similar health policies. In veterinary surveillance in Switzerland some regions are also aggregated in cantons as the cantonal veterinary authorities are responsible for the implementation of federal veterinary legislation (Schrödle et al., 2011). When there exist two or more sensible groupings, models incorporating different levels of spatial aggregations are adequate. These models permit both to identify regional effects at each level of spatial aggregation and to model spatio-temporal interactions at different levels. In addition, second level areas may have an interest *per se* to analyze for example, differences in the implementation of health policies. These potential differences may be masked when using a single level area model. In the real example analyzed in this chapter, the second-level interaction term shows differences among health areas. While some of them have an increasing effect on the global risk, others show an opposite effect (see Figure 3.3).

In this chapter several spatio-temporal models with two-levels of spatial aggregation have been proposed but multilevel models can be also constructed if needed. Simulation results have shown how these two-level models are better candidates than one single level models in different scenarios.

Some of the proposed models could be difficult to fit in INLA if the number of SLAs is too big, like Models E and F (see Section 3.2) because the computation of the posterior marginals of the hyperparameters is not straightforward given the high cost of evaluating the approximation to the joint density of hyperparameters (Martins et al., 2013). One possibility to overcome this difficulty is to substitute the LCAR prior by an iCAR prior with just a single hyperparameter. This would be appropriate when analyzing data with strong spatial dependence among the areas. The computational burden is also highly reduced if SLA instead of FLA interactions are used because the number of constraints needed for model identifiability is much smaller. In any case, if the number of SLAs is small or moderate, Models E and F could be also a useful and sensible alternative to fit in INLA. Further research on how to speed up these fitting techniques when the number of precision parameters is big is needed.

Brain cancer mortality data in the municipalities of Navarre and the Basque Country have been analyzed to illustrate the new model proposals. In this real data analysis a two-level model with SLAs interaction has overcome the single-level spatial models traditionally used in disease mapping. Although the difference in DIC is moderate, the computational time is much smaller and model interpretation is

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

richer because the second level effect is captured. The real data analysis shows that brain cancer mortality is still increasing in the municipalities of Navarre and the Basque country. Risks are significantly high in several municipalities of the health areas of Pamplona and Estella, including the more populated municipalities of these health areas also called Pamplona and Estella. Risk factors affecting the significant high risk in these areas still remain unknown.

The contents of this chapter have been published in Statistical Methods in Medical Research.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Appendix 3A: Identifiability constraints for two-level structure models

Appendix 3A shows how the sum-to-zero identifiability constraints for the two-level spatially structure models (summarized in Table 3.2) are derived.

Two-level model A

The *two-level model A* described in Equation (3.1) can be expressed in matrix form as

$$\log \mathbf{r} = (\mathbf{1}_n \otimes \mathbf{1}_T)\eta + (\mathbf{I}_n \otimes \mathbf{1}_T)\boldsymbol{\xi} + (\mathbf{M} \otimes \mathbf{1}_T)\boldsymbol{\psi} + (\mathbf{1}_n \otimes \mathbf{I}_T)\boldsymbol{\gamma} + (\mathbf{I}_n \otimes \mathbf{I}_T)\boldsymbol{\delta}, \quad (3.7)$$

with

$$\mathbf{M} = \begin{pmatrix} \mathbf{1}_{n_1} & & \\ & \ddots & \\ & & \mathbf{1}_{n_m} \end{pmatrix} \quad \text{and} \quad \sum_{j=1}^{m} n_j = n,$$

where $\mathbf{r} = (r_{11}, \ldots, r_{1T}, \ldots, r_{n1}, \ldots, r_{nT})'$, $\mathbf{1}_n$ and $\mathbf{1}_T$ are vectors of ones of length n and T respectively, $\mathbf{I}_n$ and $\mathbf{I}_T$ are $n \times n$ and $T \times T$ identity matrices respectively, and n_j denotes the number of FLAs within the jth SLA, for $j = 1, \ldots, m$. In Section 3.2, the following prior distributions are assumed for the random effects

$$\begin{aligned} \boldsymbol{\xi} &\sim N(\mathbf{0}, [\tau_\xi \mathbf{D}_\xi]^{-1}); \quad \mathbf{D}_\xi = \lambda_\xi \mathbf{R}_\xi + (1 - \lambda_\xi)\mathbf{I}_n, \\ \boldsymbol{\psi} &\sim N(\mathbf{0}, [\tau_\psi \mathbf{D}_\psi]^{-1}); \quad \mathbf{D}_\psi = \lambda_\psi \mathbf{R}_\psi + (1 - \lambda_\psi)\mathbf{I}_m, \\ \gamma &\sim N(\mathbf{0}, [\tau_\gamma \mathbf{R}_\gamma]^{-}); \quad \mathbf{R}_\gamma \equiv \text{RW1 structure matrix}, \\ \boldsymbol{\delta} &\sim N(\mathbf{0}, [\tau_\delta (\mathbf{R}_\xi \otimes \mathbf{R}_\gamma)]^{-1}); \quad \text{(Type IV interaction)}. \end{aligned}$$

Note that although all the eigenvalues of the structure matrix $\mathbf{D}_\xi$ are positive whenever $0 \leq \lambda_\xi < 1$ (i.e., $\mathbf{D}_\xi$ is a full rank matrix), the following spectral decomposition can be considered (Goicoa et al., 2017)

$$\mathbf{D}_\xi = \mathbf{U}_\xi (\lambda_\xi \boldsymbol{\Sigma}_\xi + (1 - \lambda_\xi)\mathbf{I}_n)\mathbf{U}_\xi'$$

where $\mathbf{U}_\xi = [\mathbf{U}_{\xi r} : \mathbf{U}_{\xi s}]$ is an orthogonal matrix with columns the eigenvectors of $\mathbf{R}_\xi$, $\mathbf{U}_{\xi r} = \mathbf{1}_n$ (up to a normalizing constant) and $\mathbf{U}_{\xi s}$ are the matrices of eigenvectors having null and non-null eigenvalues respectively, and $\boldsymbol{\Sigma}_\xi$ is a diagonal matrix with the eigenvalues of $\mathbf{R}_\xi$ in the main diagonal. That is, $\mathbf{D}_\xi$ has the same eigenvectors as $\mathbf{R}_\xi$ but different eigenvalues.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Since $\mathbf{U}_\xi$ is an orthogonal matrix, using the reparameterization described in Equation (1.5) we obtain that

$$\begin{aligned}(\mathbf{I}_n \otimes \mathbf{1}_T)\boldsymbol{\xi} &= (\mathbf{I}_n \otimes \mathbf{1}_T)\mathbf{U}_\xi\mathbf{U}_\xi'\boldsymbol{\xi} = (\mathbf{I}_n \otimes \mathbf{1}_T)[\mathbf{1}_n : \mathbf{U}_{\xi s}]\begin{bmatrix}\mathbf{1}_n' \\ \mathbf{U}_{\xi s}'\end{bmatrix}\boldsymbol{\xi} \\ &= (\mathbf{1}_n \otimes \mathbf{1}_T)\beta_\xi + (\mathbf{U}_{\xi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_\xi.\end{aligned}$$

In a similar way, the spectral decomposition of the structure matrices $\mathbf{D}_\psi$, $\mathbf{R}_\gamma$ and $\mathbf{R}_\delta = \mathbf{R}_\xi \otimes \mathbf{R}_\gamma$ are expressed as

$$\begin{aligned}\mathbf{D}_\psi &= \mathbf{U}_\psi(\lambda_\psi\boldsymbol{\Sigma}_\psi + (1-\lambda_\psi)\mathbf{I}_m)\mathbf{U}_\psi', \\ \mathbf{R}_\gamma &= \mathbf{U}_\gamma\boldsymbol{\Sigma}_\gamma\mathbf{U}_\gamma', \\ \mathbf{R}_\delta &= \mathbf{U}_\delta(\boldsymbol{\Sigma}_\xi \otimes \boldsymbol{\Sigma}_\gamma)\mathbf{U}_\delta',\end{aligned}$$

with

$$\mathbf{U}_\psi = [\mathbf{1}_m : \mathbf{U}_{\psi s}], \quad \mathbf{U}_\gamma = [\mathbf{1}_T \mid \mathbf{U}_{\gamma s}],$$
$$\mathbf{U}_\delta = [\mathbf{1}_n \otimes \mathbf{1}_T : \mathbf{1}_n \otimes \mathbf{U}_{\gamma s} : \mathbf{U}_{\xi s} \otimes \mathbf{1}_T \mid \mathbf{U}_{\xi s} \otimes \mathbf{U}_{\gamma s}],$$

where $\mathbf{1}_m$ is a vector of ones (up to a constant) of length m, $\mathbf{U}_{\psi s}$ is the $m \times (m-1)$ matrix whose columns are eigenvectors of $\mathbf{R}_\psi$ having non-null eigenvalues, $\mathbf{U}_{\gamma s}$ is the $T \times (T-1)$ matrix whose columns are eigenvectors of $\mathbf{R}_\gamma$ having non-null eigenvalues, $\boldsymbol{\Sigma}_\psi$ is a diagonal matrix with the eigenvalues of $\mathbf{R}_\psi$ in the main diagonal and $\boldsymbol{\Sigma}_\gamma$ is a diagonal matrix with the eigenvalues of $\mathbf{R}_\gamma$ in the main diagonal.

Since $\mathbf{U}_\psi$, $\mathbf{U}_\gamma$ and $\mathbf{U}_\delta$ are orthogonal matrices, using the reparameterization described in Equation (1.5) we obtain that

$$\begin{aligned}(\mathbf{M} \otimes \mathbf{1}_T)\boldsymbol{\psi} &= (\mathbf{M} \otimes \mathbf{1}_T)\mathbf{U}_\psi\mathbf{U}_\psi'\boldsymbol{\psi} = (\mathbf{M} \otimes \mathbf{1}_T)[\mathbf{1}_m : \mathbf{U}_{\psi s}]\begin{bmatrix}\mathbf{1}_m' \\ \mathbf{U}_{\psi s}'\end{bmatrix}\boldsymbol{\psi} \\ &= (\mathbf{M}\mathbf{1}_m \otimes \mathbf{1}_T)\beta_\psi + (\mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_\psi, \\ (\mathbf{1}_n \otimes \mathbf{I}_T)\gamma &= (\mathbf{1}_n \otimes \mathbf{I}_T)\mathbf{U}_\gamma\mathbf{U}_\gamma'\boldsymbol{\gamma} = (\mathbf{1}_n \otimes \mathbf{I}_T)[\mathbf{1}_T : \mathbf{U}_{\gamma s}]\begin{bmatrix}\mathbf{1}_T' \\ \mathbf{U}_{\gamma s}'\end{bmatrix}\boldsymbol{\gamma} \\ &= (\mathbf{1}_n \otimes \mathbf{1}_T)\beta_\gamma + (\mathbf{1}_n \otimes \mathbf{U}_{\gamma s})\boldsymbol{\alpha}_\gamma, \\ (\mathbf{I}_n \otimes \mathbf{I}_T)\boldsymbol{\delta} &= (\mathbf{I}_n \otimes \mathbf{I}_T)\mathbf{U}_\delta\mathbf{U}_\delta'\boldsymbol{\delta} \\ &= (\mathbf{I}_n \otimes \mathbf{I}_T)[\mathbf{1}_n \otimes \mathbf{1}_T : \mathbf{1}_n \otimes \mathbf{U}_{\gamma s} : \mathbf{U}_{\xi s} \otimes \mathbf{1}_T \mid \mathbf{U}_{\xi s} \otimes \mathbf{U}_{\gamma s}]\begin{bmatrix}\mathbf{1}_n' \otimes \mathbf{1}_T' \\ \mathbf{1}_n' \otimes \mathbf{U}_{\gamma s}' \\ \mathbf{U}_{\xi s}' \otimes \mathbf{1}_T' \\ \mathbf{U}_{\xi s}' \otimes \mathbf{U}_{\gamma s}'\end{bmatrix}\boldsymbol{\delta} \\ &= [\mathbf{1}_n \otimes \mathbf{1}_T : \mathbf{1}_n \otimes \mathbf{U}_{\gamma s} : \mathbf{U}_{\xi s} \otimes \mathbf{1}_T]\boldsymbol{\beta}_\delta + (\mathbf{U}_{\xi s} \otimes \mathbf{U}_{\gamma s})\boldsymbol{\alpha}_\delta.\end{aligned}$$

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Replacing these expressions into Equation (3.7), note that $\mathbf{M}\mathbf{1}_m = \mathbf{1}_n$, the two-level model A can be reformulated as

$$\begin{aligned}\log \mathbf{r} = &(\mathbf{1}_n \otimes \mathbf{1}_T)\eta + (\mathbf{1}_n \otimes \mathbf{1}_T)\beta_\xi + (\mathbf{U}_{\xi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_\xi \\ &+ (\mathbf{1}_n \otimes \mathbf{1}_T)\beta_\psi + (\mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_\psi \\ &+ (\mathbf{1}_n \otimes \mathbf{1}_T)\beta_\gamma + (\mathbf{1}_n \otimes \mathbf{U}_{\gamma s})\boldsymbol{\alpha}_\gamma \\ &+ [\mathbf{1}_n \otimes \mathbf{1}_T : \mathbf{1}_n \otimes \mathbf{U}_{\gamma s} : \mathbf{U}_{\xi s} \otimes \mathbf{1}_T]\boldsymbol{\beta}_\delta + (\mathbf{U}_{\xi s} \otimes \mathbf{U}_{\gamma s})\boldsymbol{\alpha}_\delta,\end{aligned}$$

with

$$\boldsymbol{\alpha}_\xi \sim N(\mathbf{0}, [\tau_\xi(\lambda_\xi \widetilde{\boldsymbol{\Sigma}}_\xi + (1-\lambda_\xi)\mathbf{I}_{n-1})]^{-1}),$$

$$\boldsymbol{\alpha}_\psi \sim N(\mathbf{0}, [\tau_\psi(\lambda_\psi \widetilde{\boldsymbol{\Sigma}}_\psi + (1-\lambda_\psi)\mathbf{I}_{m-1})]^{-1}),$$

$$\boldsymbol{\alpha}_\gamma \sim N(\mathbf{0}, [\tau_\gamma \widetilde{\boldsymbol{\Sigma}}_\gamma]^{-1}) \quad \text{and} \quad \boldsymbol{\alpha}_\delta \sim N(\mathbf{0}, [\tau_\delta(\widetilde{\boldsymbol{\Sigma}}_\xi \otimes \widetilde{\boldsymbol{\Sigma}}_\gamma)]^{-1}),$$

where $\widetilde{\boldsymbol{\Sigma}}_\xi$, $\widetilde{\boldsymbol{\Sigma}}_\psi$ and $\widetilde{\boldsymbol{\Sigma}}_\gamma$ are diagonal matrices with the non-null eigenvalues of $\mathbf{R}_\xi$, $\mathbf{R}_\psi$ and $\mathbf{R}_\gamma$ respectively. If we remove the repeated columns $\mathbf{1}_n \otimes \mathbf{1}_T$ (corresponding to β_ξ, β_ψ, β_γ and $\boldsymbol{\beta}_\delta$), $\mathbf{1}_n \otimes \mathbf{U}_{\gamma s}$ and $\mathbf{U}_{\xi s} \otimes \mathbf{1}_T$ (corresponding to $\boldsymbol{\beta}_\delta$), this leaves the following model

$$\begin{aligned}\log \mathbf{r} = &(\mathbf{1}_n \otimes \mathbf{1}_T)\eta + (\mathbf{U}_{\xi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_\xi + (\mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_\psi \\ &+ (\mathbf{1}_n \otimes \mathbf{U}_{\gamma s})\boldsymbol{\alpha}_\gamma + (\mathbf{U}_{\xi s} \otimes \mathbf{U}_{\gamma s})\boldsymbol{\alpha}_\delta.\end{aligned}$$

Removing (or setting to zero) the repeated columns makes the model identifiable, and now the precision matrices of the reparameterized random effects have full rank. Note that removing the repeated columns leads to the linear constraints

$$\sum_{i=1}^{n} \xi_i = 0; \quad \sum_{j=1}^{m} \psi_j = 0; \quad \sum_{t=1}^{T} \gamma_t = 0; \quad \begin{array}{ll} \sum_{t=1}^{T} \delta_{it} = 0, & \text{for} \quad i = 1, \ldots, n; \\ \sum_{i=1}^{n} \delta_{it} = 0, & \text{for} \quad t = 1, \ldots, T. \end{array}$$

See Goicoa et al. (2017) for details about the derivation of these sum-to-zero constraints.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Two-level model B

The *two-level model B* described in Equation (3.2) can be expressed in matrix form as

$$\log \mathbf{r} = (\mathbf{1}_n \otimes \mathbf{1}_T)\eta + (\mathbf{I}_n \otimes \mathbf{1}_T)\boldsymbol{\xi} + (\mathbf{M} \otimes \mathbf{1}_T)\boldsymbol{\psi} + (\mathbf{1}_n \otimes \mathbf{I}_T)\boldsymbol{\gamma} + (\mathbf{M} \otimes \mathbf{I}_T)\boldsymbol{\delta}, \quad (3.8)$$

where now $\boldsymbol{\delta} = (\delta_{11}, \ldots, \delta_{1T}, \ldots, \delta_{m1}, \ldots, \delta_{mT})'$ represents the SLA space-time interaction and the following prior distribution is assumed

$$\boldsymbol{\delta} \sim N(\mathbf{0}, [\tau_\delta(\mathbf{R}_\psi \otimes \mathbf{R}_\gamma)]^{-1}); \quad \text{(Type IV interaction)}.$$

Consider the spectral decomposition of the structure matrix $\mathbf{R}_\delta = \mathbf{R}_\psi \otimes \mathbf{R}_\gamma$,

$$\mathbf{R}_\delta = \mathbf{U}_\delta(\boldsymbol{\Sigma}_\psi \otimes \boldsymbol{\Sigma}_\gamma)\mathbf{U}_\delta',$$

with

$$\mathbf{U}_\delta = [\mathbf{1}_m \otimes \mathbf{1}_T : \mathbf{1}_m \otimes \mathbf{U}_{\gamma s} : \mathbf{U}_{\psi s} \otimes \mathbf{1}_T \mid \mathbf{U}_{\psi s} \otimes \mathbf{U}_{\gamma s}].$$

Then, as $\mathbf{U}_\delta$ is an orthogonal matrix and $\mathbf{M}\mathbf{1}_m = \mathbf{1}_n$

$$\begin{aligned}
(\mathbf{M} \otimes \mathbf{I}_T)\boldsymbol{\delta} &= (\mathbf{M} \otimes \mathbf{I}_T)\mathbf{U}_\delta\mathbf{U}_\delta'\boldsymbol{\delta} \\
&= (\mathbf{M} \otimes \mathbf{I}_T)[\mathbf{1}_m \otimes \mathbf{1}_T : \mathbf{1}_m \otimes \mathbf{U}_{\gamma s} : \mathbf{U}_{\psi s} \otimes \mathbf{1}_T \mid \mathbf{U}_{\psi s} \otimes \mathbf{U}_{\gamma s}] \begin{bmatrix} \mathbf{1}_m' \otimes \mathbf{1}_T' \\ \mathbf{1}_m' \otimes \mathbf{U}_{\gamma s}' \\ \mathbf{U}_{\psi s}' \otimes \mathbf{1}_T' \\ \mathbf{U}_{\psi s}' \otimes \mathbf{U}_{\gamma s}' \end{bmatrix} \boldsymbol{\delta} \\
&= [\mathbf{1}_n \otimes \mathbf{1}_T : \mathbf{1}_n \otimes \mathbf{U}_{\gamma s} : \mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{1}_T]\boldsymbol{\beta}_\delta + (\mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{U}_{\gamma s})\boldsymbol{\alpha}_\delta,
\end{aligned}$$

As in the previous case, replacing this expressions into Equation (3.8) the two-level model B can be reformulated as

$$\begin{aligned}
\log \mathbf{r} = &(\mathbf{1}_n \otimes \mathbf{1}_T)\eta + (\mathbf{1}_n \otimes \mathbf{1}_T)\beta_\xi + (\mathbf{U}_{\xi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_\xi \\
&+ (\mathbf{1}_n \otimes \mathbf{1}_T)\beta_\psi + (\mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_\psi \\
&+ (\mathbf{1}_n \otimes \mathbf{1}_T)\beta_\gamma + (\mathbf{1}_n \otimes \mathbf{U}_{\gamma s})\boldsymbol{\alpha}_\gamma \\
&+ [\mathbf{1}_n \otimes \mathbf{1}_T : \mathbf{1}_n \otimes \mathbf{U}_{\gamma s} : \mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{1}_T]\boldsymbol{\beta}_\delta + (\mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{U}_{\gamma s})\boldsymbol{\alpha}_\delta,
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{\alpha}_\xi &\sim N(\mathbf{0}, [\tau_\xi(\lambda_\xi\widetilde{\boldsymbol{\Sigma}}_\xi + (1-\lambda_\xi)\mathbf{I}_{n-1})]^{-1}), \\
\boldsymbol{\alpha}_\psi &\sim N(\mathbf{0}, [\tau_\psi(\lambda_\psi\widetilde{\boldsymbol{\Sigma}}_\psi + (1-\lambda_\psi)\mathbf{I}_{m-1})]^{-1}), \\
\boldsymbol{\alpha}_\gamma \sim N(\mathbf{0}, [\tau_\gamma\widetilde{\boldsymbol{\Sigma}}_\gamma]^{-1}) \quad &\text{and} \quad \boldsymbol{\alpha}_\delta \sim N(\mathbf{0}, [\tau_\delta(\widetilde{\boldsymbol{\Sigma}}_\psi \otimes \widetilde{\boldsymbol{\Sigma}}_\gamma)]^{-1}).
\end{aligned}$$

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

If we remove the repeated columns $\mathbf{1}_n \otimes \mathbf{1}_T$ (corresponding to β_ξ, β_ψ, β_γ and $\boldsymbol{\beta}_\delta$), $\mathbf{1}_n \otimes \mathbf{U}_{\gamma s}$ and $\mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{1}_T$ (corresponding to $\boldsymbol{\beta}_\delta$), this leaves the following model

$$\log \mathbf{r} = (\mathbf{1}_n \otimes \mathbf{1}_T)\eta + (\mathbf{U}_{\xi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_\xi + (\mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_\psi + (\mathbf{1}_n \otimes \mathbf{U}_{\gamma s})\boldsymbol{\alpha}_\gamma + (\mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{U}_{\gamma s})\boldsymbol{\alpha}_\delta.$$

Removing (or setting to zero) the repeated columns makes the model identifiable, and now the precision matrices of the reparameterized random effects have full rank. Note that removing the repeated columns leads to the linear constraints

$$\sum_{i=1}^{n} \xi_i = 0; \quad \sum_{j=1}^{m} \psi_j = 0; \quad \sum_{t=1}^{T} \gamma_t = 0; \quad \begin{array}{ll} \sum_{t=1}^{T} \delta_{jt} = 0, & \text{for} \quad j = 1, \ldots, m; \\ \sum_{j=1}^{m} \delta_{jt} = 0, & \text{for} \quad t = 1, \ldots, T. \end{array}$$

Two-level model C

The *two-level model C* described in Equation (3.4) can be expressed in matrix form as

$$\log \mathbf{r} = (\mathbf{1}_n \otimes \mathbf{1}_T)\eta + (\mathbf{I}_n \otimes \mathbf{1}_T)\boldsymbol{\xi}^* + (\mathbf{M} \otimes \mathbf{1}_T)\boldsymbol{\psi} + (\mathbf{1}_n \otimes \mathbf{I}_T)\boldsymbol{\gamma} + (\mathbf{I}_n \otimes \mathbf{I}_T)\boldsymbol{\delta}^*, \quad (3.9)$$

and the following prior distributions are assumed for the random effects

$$\begin{aligned} \boldsymbol{\xi}^* &\sim N(\mathbf{0}, [\tau_\xi \mathbf{D}_\xi^*]^{-1}); \quad \mathbf{D}_\xi^* = \lambda_\xi \mathbf{R}_\xi^* + (1-\lambda_\xi)\mathbf{I}_n, \\ \boldsymbol{\psi} &\sim N(\mathbf{0}, [\tau_\psi \mathbf{D}_\psi]^{-1}); \quad \mathbf{D}_\psi = \lambda_\psi \mathbf{R}_\psi + (1-\lambda_\psi)\mathbf{I}_m, \\ \gamma &\sim N(\mathbf{0}, [\tau_\gamma \mathbf{R}_\gamma]^{-}); \quad \mathbf{R}_\gamma \equiv \text{RW1 structure matrix}, \\ \boldsymbol{\delta}^* &\sim N(\mathbf{0}, [\tau_\delta (\mathbf{R}_\xi^* \otimes \mathbf{R}_\gamma)]^{-1}); \quad \text{(Type IV interaction)}. \end{aligned}$$

Recall that $\mathbf{R}_\xi^* = \text{blockdiag}(\mathbf{R}_{\xi_1}, \ldots, \mathbf{R}_{\xi_m})$ is a block-diagonal matrix where $\mathbf{R}_{\xi_j}$ is the neighborhood matrix of FLAs within the jth SLA, for $j = 1, \ldots, m$. Again, although all the eigenvalues of the matrix $\mathbf{D}_\xi^*$ are positive whenever $0 \leq \lambda_\xi < 1$ (i.e., $\mathbf{D}_\xi^*$ is a full rank matrix), the following spectral decomposition can be considered

$$\mathbf{D}_\xi^* = \mathbf{U}_\xi^* (\lambda_\xi \boldsymbol{\Sigma}_\xi^* + (1-\lambda_\xi)\mathbf{I}_n)\mathbf{U}_\xi^{*'},$$

where $\mathbf{U}_\xi^* = [\mathbf{U}_{\xi r}^* : \mathbf{U}_{\xi s}^*]$ is an orthogonal matrix with columns the eigenvectors of

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

$\mathbf{R}^*_\xi$, and $\mathbf{\Sigma}^*_\xi$ is a diagonal matrix with the eigenvalues of $\mathbf{R}^*_\xi$ in the main diagonal. That is, $\mathbf{D}^*_\xi$ has the same eigenvectors as $\mathbf{R}^*_\xi$ but different eigenvalues. Note that in fact, the columns of $\mathbf{M}$ are the null eigenvalues of $\mathbf{R}^*_\xi$, so we define $\mathbf{U}^*_{\xi r} = \mathbf{M}$ (up to a normalizing constant) and $\mathbf{U}^*_{\xi s}$ is a $n \times (n-m)$ matrix with columns the non-null eigenvectors of $\mathbf{R}^*_\xi$.

Since $\mathbf{U}^*_\xi$ is an orthogonal matrix, using the reparameterization described in Equation (1.5) we obtain that

$$\begin{aligned}(\mathbf{I}_n \otimes \mathbf{1}_T)\boldsymbol{\xi}^* &= (\mathbf{I}_n \otimes \mathbf{1}_T)\mathbf{U}^*_\xi \mathbf{U}^{*'}_\xi \boldsymbol{\xi}^* = (\mathbf{I}_n \otimes \mathbf{1}_T)[\mathbf{M} : \mathbf{U}^*_{\xi s}] \begin{bmatrix} \mathbf{M}' \\ \mathbf{U}^{*'}_{\xi s} \end{bmatrix} \boldsymbol{\xi}^* \\ &= (\mathbf{M} \otimes \mathbf{1}_T)\boldsymbol{\beta}_{\xi^*} + (\mathbf{U}^*_{\xi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_{\xi^*}.\end{aligned}$$

Similarly, the spectral decomposition of the structure matrix $\mathbf{R}^*_\delta = \mathbf{R}^*_\xi \otimes \mathbf{R}_\gamma$ is expressed as

$$\mathbf{R}^*_\delta = \mathbf{U}^*_\delta(\mathbf{\Sigma}^*_\xi \otimes \mathbf{\Sigma}_\gamma)\mathbf{U}^{*'}_\delta,$$

with

$$\mathbf{U}^*_\delta = [\mathbf{M} \otimes \mathbf{1}_T : \mathbf{M} \otimes \mathbf{U}_{\gamma s} : \mathbf{U}^*_{\xi s} \otimes \mathbf{1}_T \mid \mathbf{U}^*_{\xi s} \otimes \mathbf{U}_{\gamma s}].$$

Then, as $\mathbf{U}^*_\delta$ is an orthogonal matrix,

$$\begin{aligned}(\mathbf{I}_n \otimes \mathbf{I}_T)\boldsymbol{\delta}^* &= (\mathbf{I}_n \otimes \mathbf{I}_T)\mathbf{U}^*_\delta \mathbf{U}^{*'}_\delta \boldsymbol{\delta}^* \\ &= (\mathbf{I}_n \otimes \mathbf{I}_T)[\mathbf{M} \otimes \mathbf{1}_T : \mathbf{M} \otimes \mathbf{U}_{\gamma s} : \mathbf{U}^*_{\xi s} \otimes \mathbf{1}_T \mid \mathbf{U}^*_{\xi s} \otimes \mathbf{U}_{\gamma s}] \begin{bmatrix} \mathbf{M}' \otimes \mathbf{1}'_T \\ \mathbf{M}' \otimes \mathbf{U}'_{\gamma s} \\ \mathbf{U}^{*'}_{\xi s} \otimes \mathbf{1}'_T \\ \mathbf{U}^{*'}_{\xi s} \otimes \mathbf{U}'_{\gamma s} \end{bmatrix} \boldsymbol{\delta}^* \\ &= [\mathbf{M} \otimes \mathbf{1}_T : \mathbf{M} \otimes \mathbf{U}_{\gamma s} : \mathbf{U}^*_{\xi s} \otimes \mathbf{1}_T]\boldsymbol{\beta}_{\delta^*} + (\mathbf{U}^*_{\xi s} \otimes \mathbf{U}_{\gamma s})\boldsymbol{\alpha}_{\delta^*}.\end{aligned}$$

Replacing these expressions into Equation (3.9), the two-level model C can be reformulated as

$$\begin{aligned}\log \mathbf{r} = &(\mathbf{1}_n \otimes \mathbf{1}_T)\eta + (\mathbf{M} \otimes \mathbf{1}_T)\boldsymbol{\beta}_{\xi^*} + (\mathbf{U}^*_{\xi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_{\xi^*} \\ &+ (\mathbf{1}_n \otimes \mathbf{1}_T)\beta_\psi + (\mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_\psi \\ &+ (\mathbf{1}_n \otimes \mathbf{1}_T)\beta_\gamma + (\mathbf{1}_n \otimes \mathbf{U}_{\gamma s})\boldsymbol{\alpha}_\gamma \\ &+ [\mathbf{M} \otimes \mathbf{1}_T : \mathbf{M} \otimes \mathbf{U}_{\gamma s} : \mathbf{U}^*_{\xi s} \otimes \mathbf{1}_T]\boldsymbol{\beta}_{\delta^*} + (\mathbf{U}^*_{\xi s} \otimes \mathbf{U}_{\gamma s})\boldsymbol{\alpha}_{\delta^*}\end{aligned}$$

with

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

$$\boldsymbol{\alpha}_{\xi^*} \sim N(\mathbf{0}, [\tau_\xi(\lambda_\xi \widetilde{\boldsymbol{\Sigma}}_\xi^* + (1-\lambda_\xi)\mathbf{I}_{n-m})]^{-1}),$$

$$\boldsymbol{\alpha}_{\psi} \sim N(\mathbf{0}, [\tau_\psi(\lambda_\psi \widetilde{\boldsymbol{\Sigma}}_\psi + (1-\lambda_\psi)\mathbf{I}_{m-1})]^{-1}),$$

$$\boldsymbol{\alpha}_{\gamma} \sim N(\mathbf{0}, [\tau_\gamma \widetilde{\boldsymbol{\Sigma}}_\gamma]^{-1}) \quad \text{and} \quad \boldsymbol{\alpha}_{\delta^*} \sim N(\mathbf{0}, [\tau_\delta(\widetilde{\boldsymbol{\Sigma}}_\xi^* \otimes \widetilde{\boldsymbol{\Sigma}}_\gamma)]^{-1}),$$

where $\widetilde{\boldsymbol{\Sigma}}_\xi^*$, $\widetilde{\boldsymbol{\Sigma}}_\psi$ and $\widetilde{\boldsymbol{\Sigma}}_\gamma$ are diagonal matrices with the non-null eigenvalues of $\mathbf{R}_\xi^*$, $\mathbf{R}_\psi$ and $\mathbf{R}_\gamma$ respectively. If we remove the repeated columns $\mathbf{1}_n \otimes \mathbf{1}_T$ (corresponding to β_ψ and β_γ), $\mathbf{U}_{\xi s}^* \otimes \mathbf{1}_T$ (corresponding to $\boldsymbol{\beta}_\delta$), and since the columns of $\mathbf{M} \otimes \mathbf{1}_T$ are linear combinations of $\mathbf{1}_n \otimes \mathbf{1}_t$ and $\mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{1}_T$ (corresponding to $\boldsymbol{\beta}_{\xi^*}$ and $\boldsymbol{\beta}_{\delta^*}$), this leaves the following model

$$\begin{aligned}\log \mathbf{r} = {} & (\mathbf{1}_n \otimes \mathbf{1}_T)\eta + (\mathbf{U}_{\xi s}^* \otimes \mathbf{1}_T)\boldsymbol{\alpha}_{\xi^*} + (\mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_{\psi} \\ & + (\mathbf{1}_n \otimes \mathbf{U}_{\gamma s})\boldsymbol{\alpha}_{\gamma} + (\mathbf{U}_{\xi s}^* \otimes \mathbf{U}_{\gamma s})\boldsymbol{\alpha}_{\delta^*}.\end{aligned}$$

Removing (or setting to zero) the repeated columns makes the model identifiable, and now the precision matrices of the reparameterized random effects have full rank. Note that removing the repeated columns leads to the linear constraints

$$\sum_{i=1}^{n_j} \xi_{j,i} = 0, \quad \text{for} \quad j = 1, \ldots, m; \quad \sum_{j=1}^{m} \psi_j = 0; \quad \sum_{t=1}^{T} \gamma_t = 0;$$
$$\sum_{t=1}^{T} \delta_{it} = 0, \quad \text{for} \quad i = 1, \ldots, n;$$
$$\sum_{i=1}^{n_j} \delta_{j,it} = 0, \quad \text{for} \quad \begin{matrix} j = 1, \ldots, m; \\ t = 1, \ldots, T. \end{matrix}$$

Two-level model D

The *two-level model D* described in Equation (3.5) can be expressed in matrix form as

$$\log \mathbf{r} = (\mathbf{1}_n \otimes \mathbf{1}_T)\eta + (\mathbf{I}_n \otimes \mathbf{1}_T)\boldsymbol{\xi}^* + (\mathbf{M} \otimes \mathbf{1}_T)\boldsymbol{\psi} + (\mathbf{1}_n \otimes \mathbf{I}_T)\boldsymbol{\gamma} + (\mathbf{M} \otimes \mathbf{I}_T)\boldsymbol{\delta}, \quad (3.10)$$

where as in two-level model B, $\boldsymbol{\delta} = (\delta_{11}, \ldots, \delta_{1T}, \ldots, \delta_{m1}, \ldots, \delta_{mT})'$ represents the SLA space-time interaction and the following prior distribution is assumed

$$\boldsymbol{\delta} \sim N(\mathbf{0}, [\tau_\delta(\mathbf{R}_\psi \otimes \mathbf{R}_\gamma)]^{-1}); \quad \text{(Type IV interaction)}.$$

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Considering the spectral decomposition of the structure matrices $\mathbf{R}^*_\xi$, $\mathbf{R}_\psi$, $\mathbf{R}_\gamma$ and $\mathbf{R}_\delta = \mathbf{R}_\psi \otimes \mathbf{R}_\gamma$ described before, the two-level model D can be reformulated as

$$\begin{aligned}\log \mathbf{r} = &(\mathbf{1}_n \otimes \mathbf{1}_T)\eta + (\mathbf{M} \otimes \mathbf{1}_T)\boldsymbol{\beta}_{\xi^*} + (\mathbf{U}^*_{\xi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_{\xi^*} \\ &+ (\mathbf{1}_n \otimes \mathbf{1}_T)\beta_\psi + (\mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_\psi \\ &+ (\mathbf{1}_n \otimes \mathbf{1}_T)\beta_\gamma + (\mathbf{1}_n \otimes \mathbf{U}_{\gamma s})\boldsymbol{\alpha}_\gamma \\ &+ [\mathbf{1}_n \otimes \mathbf{1}_T : \mathbf{1}_n \otimes \mathbf{U}_{\gamma s} : \mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{1}_T]\boldsymbol{\beta}_\delta + (\mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{U}_{\gamma s})\boldsymbol{\alpha}_\delta\end{aligned}$$

with

$$\begin{aligned}\boldsymbol{\alpha}_{\xi^*} &\sim N(\mathbf{0}, [\tau_\xi(\lambda_\xi \widetilde{\boldsymbol{\Sigma}}^*_\xi + (1-\lambda_\xi)\mathbf{I}_{n-m})]^{-1}), \\ \boldsymbol{\alpha}_\psi &\sim N(\mathbf{0}, [\tau_\psi(\lambda_\psi \widetilde{\boldsymbol{\Sigma}}_\psi + (1-\lambda_\psi)\mathbf{I}_{m-1})]^{-1}), \\ \boldsymbol{\alpha}_\gamma \sim N(\mathbf{0}, [\tau_\gamma \widetilde{\boldsymbol{\Sigma}}_\gamma]^{-1}) \quad &\text{and} \quad \boldsymbol{\alpha}_\delta \sim N(\mathbf{0}, [\tau_\delta(\widetilde{\boldsymbol{\Sigma}}_\psi \otimes \widetilde{\boldsymbol{\Sigma}}_\gamma)]^{-1}).\end{aligned}$$

If we remove the repeated columns $\mathbf{1}_n \otimes \mathbf{1}_T$ (corresponding to β_ψ, β_γ and $\boldsymbol{\beta}_\delta$), $\mathbf{1}_n \otimes \mathbf{U}_{\gamma s}$ and $\mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{1}_T$ (corresponding to $\boldsymbol{\beta}_\delta$), and since the columns of $\mathbf{M} \otimes \mathbf{1}_T$ are linear combinations of $\mathbf{1}_n \otimes \mathbf{1}_T$ and $\mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{1}_T$ (corresponding to this $\boldsymbol{\beta}^*_\xi$), this leaves the following model

$$\begin{aligned}\log \mathbf{r} = &(\mathbf{1}_n \otimes \mathbf{1}_T)\eta + (\mathbf{U}^*_{\xi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_{\xi^*} + (\mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_\psi \\ &+ (\mathbf{1}_n \otimes \mathbf{U}_{\gamma s})\boldsymbol{\alpha}_\gamma + (\mathbf{M}\mathbf{U}_{\psi s} \otimes \mathbf{U}_{\gamma s})\boldsymbol{\alpha}_\delta.\end{aligned}$$

Removing (or setting to zero) the repeated columns makes the model identifiable, and now the precision matrices of the reparameterized random effects have full rank. Note that removing the repeated columns leads to the linear constraints

$$\sum_{i=1}^{n_j} \xi_{j,i} = 0, \quad \text{for} \quad j = 1, \ldots, m; \qquad \sum_{j=1}^{m} \psi_j = 0; \qquad \sum_{t=1}^{T} \gamma_t = 0;$$
$$\sum_{t=1}^{T} \delta_{jt} = 0, \quad \text{for} \quad j = 1, \ldots, m;$$
$$\sum_{j=1}^{m} \delta_{jt} = 0, \quad \text{for} \quad t = 1, \ldots, T.$$

Two-level models E and F

Further extensions of two-level models C and D (named *two-level model E* for FLA

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

interactions and *two-level model F* for SLA interactions) are proposed in Section 3.2 including m independent random effects to model the spatial dependence within each SLA.

In these models, different precision components are attached to the blocks of the structure matrix $\mathbf{D}_{\xi}^{*}$, so that the spatial random effect is decomposed as $\boldsymbol{\xi}^{*} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m)'$ and the following prior distributions are assumed

$$\boldsymbol{\xi}_j \sim N(\mathbf{0}, [\tau_{\xi_j}\mathbf{D}_{\xi_j}]^{-1}); \quad \mathbf{D}_{\xi_j} = \lambda_{\xi_j}\mathbf{R}_{\xi_j} + (1-\lambda_{\xi_j})\mathbf{I}_{n_j}, \quad j = 1, \ldots, m.$$

As for the previous models, although all the eigenvalues of the matrices $\mathbf{D}_{\xi_j}$ are positive whenever $0 \leq \lambda_{\xi_j} < 1$ (i.e., $\mathbf{D}_{\xi_j}$ are full rank matrices), the following spectral decompositions can be considered

$$\mathbf{D}_{\xi_j} = \mathbf{U}_{\xi_j}(\lambda_{\xi_j}\boldsymbol{\Sigma}_{\xi_j} + (1-\lambda_{\xi_j})\mathbf{I}_{n_j})\mathbf{U}_{\xi_j}',$$

where $\mathbf{U}_{\xi_j} = [\mathbf{U}_{\xi_j r} : \mathbf{U}_{\xi_j s}]$ are orthogonal matrices with columns the eigenvectors of $\mathbf{R}_{\xi_j}$, $\mathbf{U}_{\xi_j r} = \mathbf{1}_{n_j}$ (up to a normalizing constant) and $\mathbf{U}_{\xi_j s}$ are the matrices of eigenvectors having null and non-null eigenvalues respectively, and $\boldsymbol{\Sigma}_{\xi_j}$ are diagonal matrices with the eigenvalues of $\mathbf{R}_{\xi_j}$ in the main diagonal. That is, each $\mathbf{D}_{\xi_j}$ has the same eigenvector as $\mathbf{R}_{\xi_j}$ but different eigenvalues.

Since each $\mathbf{U}_{\xi_j}$ is an orthogonal matrix, using the reparameterization described in Equation (1.5) we obtain

$$\begin{aligned}
(\mathbf{I}_n \otimes \mathbf{1}_T)\boldsymbol{\xi}^{*} &= \left[\begin{pmatrix} \mathbf{I}_{n_1} & & \\ & \ddots & \\ & & \mathbf{I}_{n_m} \end{pmatrix} \otimes \mathbf{1}_T\right] \begin{bmatrix} \boldsymbol{\xi}_1 \\ \vdots \\ \boldsymbol{\xi}_m \end{bmatrix} \\
&= \begin{pmatrix} \mathbf{I}_{n_1}\boldsymbol{\xi}_1 & & \\ & \ddots & \\ & & \mathbf{I}_{n_m}\boldsymbol{\xi}_m \end{pmatrix} \otimes \mathbf{1}_T \\
&= \begin{pmatrix} \mathbf{I}_{n_1}\mathbf{U}_{\xi_1}\mathbf{U}_{\xi_1}'\boldsymbol{\xi}_1 & & \\ & \ddots & \\ & & \mathbf{I}_{n_m}\mathbf{U}_{\xi_m}\mathbf{U}_{\xi_m}'\boldsymbol{\xi}_m \end{pmatrix} \otimes \mathbf{1}_T \\
&= \begin{pmatrix} \mathbf{I}_{n_1}[\mathbf{1}_{n_1} : \mathbf{U}_{\xi_1 s}] \begin{bmatrix} \mathbf{1}_{n_1}'\boldsymbol{\xi}_1 \\ \mathbf{U}_{\xi_1 s}'\boldsymbol{\xi}_1 \end{bmatrix} & & \\ & \ddots & \\ & & \mathbf{I}_{n_m}[\mathbf{1}_{n_m} : \mathbf{U}_{\xi_m s}] \begin{bmatrix} \mathbf{1}_{n_m}'\boldsymbol{\xi}_m \\ \mathbf{U}_{\xi_m s}'\boldsymbol{\xi}_m \end{bmatrix} \end{pmatrix} \otimes \mathbf{1}_T
\end{aligned}$$

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

$$= \begin{pmatrix} \mathbf{1}_{n_1}\beta_{\xi_1} + \mathbf{U}_{\xi_1 s}\boldsymbol{\alpha}_{\xi_1} & & \\ & \ddots & \\ & & \mathbf{1}_{n_m}\beta_{\xi_m} + \mathbf{U}_{\xi_m s}\boldsymbol{\alpha}_{\xi_m} \end{pmatrix} \otimes \mathbf{1}_T$$

$$= \left[\begin{pmatrix} \mathbf{1}_{n_1} & & \\ & \ddots & \\ & & \mathbf{1}_{n_m} \end{pmatrix} \begin{bmatrix} \beta_{\xi_1} \\ \vdots \\ \beta_{\xi_m} \end{bmatrix} + \begin{pmatrix} \mathbf{U}_{\xi_1 s} & & \\ & \ddots & \\ & & \mathbf{U}_{\xi_m s} \end{pmatrix} \begin{bmatrix} \boldsymbol{\alpha}_{\xi_1} \\ \vdots \\ \boldsymbol{\alpha}_{\xi_m} \end{bmatrix} \right] \otimes \mathbf{1}_T$$

$$= (\mathbf{M} \otimes \mathbf{1}_T) \begin{bmatrix} \beta_{\xi_1} \\ \vdots \\ \beta_{\xi_m} \end{bmatrix} + (\mathbf{U}^*_{\xi s} \otimes \mathbf{1}_T) \begin{bmatrix} \boldsymbol{\alpha}_{\xi_1} \\ \vdots \\ \boldsymbol{\alpha}_{\xi_m} \end{bmatrix}$$

with

$$\boldsymbol{\alpha}_{\xi_j} \sim N(\mathbf{0}, [\tau_{\xi_j}(\lambda_{\xi_j}\widetilde{\boldsymbol{\Sigma}}_{\xi_j} + (1-\lambda_{\xi_j})\mathbf{I}_{n_j-1})]^{-1}), \quad \text{for} \quad j = 1, \ldots, m,$$

where $\mathbf{U}^*_{\xi s} = [\mathbf{U}_{\xi_1 s} \ \ldots \ \mathbf{U}_{\xi_m s}]$ is the $n \times (n-m)$ matrix with columns the non-null eigenvectors of $\mathbf{R}_{\xi_j}$, and each $\widetilde{\boldsymbol{\Sigma}}_{\xi_j}$ is a diagonal matrix with the non-null eigenvalues of $\mathbf{R}_{\xi_j}$. Note that the fixed effect design matrix in this expression is the same as the one obtained in the reparameterization of the FLA spatial random effect for two-level models C and D, and consequently, the same repeated columns have to be removed leading to the linear constraints

$$\boxed{\sum_{i=1}^{n_j} \xi_{j,i} = 0, \quad \text{for} \quad j = 1, \ldots, m.}$$

The same constraints as in models D and E have to be imposed for the rest of random effects ($\boldsymbol{\psi}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$).

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Appendix 3B: R code for model fitting in INLA

The R code to fit some of the two-level structure models described in Section 3.2 with INLA is detailed below. First, the data frame (or list) containing the variables of the model has to be defined, that is, the vectors with the observed and expected deaths (in this example O and E respectively) and the corresponding IDs for first-level areas (ID.FLA), second-level areas (ID.SLA), time points (ID.year), FLA space-time interaction (ID.FLA.year) and SLA space-time interaction (ID.SLA.year). Note that data must be ordered according to the Kronecker product defined for the structure matrices of the space-time interaction random effects $\boldsymbol{\delta}$ and $\boldsymbol{\delta}^*$.

Then, we define the spatial neighborhood matrices $\mathbf{R}_\xi$ and $\mathbf{R}_\psi$, as well as the required structure matrices to implement the LCAR prior distribution using the "generic1" model. Namely,

```
> g <- inla.read.graph("FLA_nb.inla")
> R.xi = matrix(0, g$n, g$n)
> for (i in 1:g$n){
+ R.xi[i,i]=g$nnbs[[i]]
+ R.xi[i,g$nbs[[i]]]=-1
+ }
> R.Leroux.FLA <- diag(n)-R.xi

> g <- inla.read.graph("SLA_nb.inla")
> R.psi = matrix(0, g$n, g$n)
> for (i in 1:g$n){
+ R.psi[i,i]=g$nnbs[[i]]
+ R.psi[i,g$nbs[[i]]]=-1
+ }
> R.Leroux.SLA <- diag(m)-R.psi
```

where "FLA_nb.inla" and "SLA_nb.inla" are respectively, the inla.graph objects containing the neighboring structures of the municipalities (n=501 first-level areas) and health areas (m=13 second-level areas) of Navarre and Basque Country. The temporal structure matrix $\mathbf{R}_\gamma$ of a random walk of first order is defined as

```
> D1 <- diff(diag(t),differences=1)
> R.gammaRW1 <- t(D1)%*%D1
```

where t is the number of time periods for which data is available (t=23 years for brain cancer mortality data).
Finally, the formula object for the models described in Table 3.1 are defined below.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Two-level model A

```
> R <- kronecker(R.xi,R.gammaRW1)
> r.def <- n+t-1
> A1 <- kronecker(diag(n),matrix(1,1,t))
> A2 <- kronecker(matrix(1,1,n),diag(t))
> delta.constr <- rbind(A1,A2)
```

```
> formula <- O ~ f(ID.FLA, model="generic1", Cmatrix=R.Leroux.FLA, constr=TRUE,
+                  hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
+                f(ID.SLA, model="generic1", Cmatrix=R.Leroux.SLA, constr=TRUE,
+                  hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
+                f(ID.year, model="rw1", hyper=list(prec=list(prior=sdunif))) +
+                f(ID.FLA.year, model="generic0", Cmatrix=R, rankdef=r.def,
+                  constr=TRUE, hyper=list(prec=list(prior=sdunif)),
+                  extraconstr=list(A=delta.constr, e=rep(0,n+t)))
```

where `R` is the structure matrix $\mathbf{R}_\xi \otimes \mathbf{R}_\gamma$ and the linear constraints that makes the model identifiable (see Table 3.2) are specified through the `constr=TRUE` and `extraconstr` arguments. Note that for the *two-level model A*, the constraints over the interaction term are expressed as

$$\sum_{t=1}^{T} \delta_{it} = 0, \quad \text{for} \quad i = 1, \ldots, n \iff (\mathbf{I}_n \otimes \mathbf{1}_T')\boldsymbol{\delta} = 0,$$
$$\sum_{i=1}^{n} \delta_{it} = 0, \quad \text{for} \quad t = 1, \ldots, T \iff (\mathbf{1}_n' \otimes \mathbf{I}_T)\boldsymbol{\delta} = 0.$$

Two-level model B

```
> R <- kronecker(R.psi,R.gammaRW1)
> r.def <- m+t-1
> A1 <- kronecker(diag(m),matrix(1,1,t))
> A2 <- kronecker(matrix(1,1,m),diag(t))
> delta.constr <- rbind(A1,A2)
```

```
> formula <- O ~ f(ID.FLA, model="generic1", Cmatrix=R.Leroux.FLA, constr=TRUE,
+                  hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
+                f(ID.SLA, model="generic1", Cmatrix=R.Leroux.SLA, constr=TRUE,
+                  hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
+                f(ID.year, model="rw1", hyper=list(prec=list(prior=sdunif))) +
+                f(ID.SLA.year, model="generic0", Cmatrix=R, rankdef=r.def,
+                  constr=TRUE, hyper=list(prec=list(prior=sdunif)),
+                  extraconstr=list(A=delta.constr, e=rep(0,m+t)))
```

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

where R is the structure matrix $\mathbf{R}_\psi \otimes \mathbf{R}_\gamma$ and the linear constraints over the interaction term for the *two-level model B* are expressed as

$$\sum_{t=1}^{T} \delta_{jt} = 0, \quad \text{for} \quad j = 1, \ldots, m \iff (\mathbf{I}_m \otimes \mathbf{1}_T')\boldsymbol{\delta} = 0,$$

$$\sum_{j=1}^{m} \delta_{jt} = 0, \quad \text{for} \quad t = 1, \ldots, T \iff (\mathbf{1}_m' \otimes \mathbf{I}_T)\boldsymbol{\delta} = 0.$$

Two-level model C

To fit the two-level models C and D, the neighborhood structure matrix $\mathbf{R}_\xi^*$ of those municipalities (FLAs) belonging to the same health area (SLA) has to be defined. Assuming that this matrix is saved in the object `R.xi.block`, the block-diagonal matrix of Equation (3.3) is defined as

```
> R.Leroux.block <- diag(n)-R.xi.block
```

Note that if data is properly ordered, the identifiability constraints for the FLA random effect $\boldsymbol{\xi}^* = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m)'$ are expressed as

$$\sum_{i=1}^{n_j} \xi_{j,i} = 0, \quad \text{for} \quad j = 1, \ldots, m \iff \begin{pmatrix} \mathbf{1}_{n_1}' & & \\ & \ddots & \\ & & \mathbf{1}_{n_m}' \end{pmatrix} \boldsymbol{\xi}^* = 0,$$

while the linear constraints for the interaction term are given by

$$\sum_{t=1}^{T} \delta_{it} = 0, \quad \text{for} \quad i = 1, \ldots, n \iff (\mathbf{I}_n \otimes \mathbf{1}_T')\boldsymbol{\delta}^* = 0,$$

$$\sum_{i=1}^{n_j} \delta_{j,it} = 0, \quad \text{for} \quad \begin{matrix} j = 1, \ldots, m. \\ t = 1, \ldots, T. \end{matrix} \iff \left[\begin{pmatrix} \mathbf{1}_{n_1}' & & \\ & \ddots & \\ & & \mathbf{1}_{n_m}' \end{pmatrix} \otimes \mathbf{I}_T \right] \boldsymbol{\delta}^* = 0.$$

So the `formula` object is defined as

```
> FLA.constr <- bdiag(matrix(1,1,N1),...,matrix(1,1,Nm))
> R <- kronecker(R.xi.block,R.gammaRW1)
> r.def <- n+m*(t-1)
> A1 <- kronecker(diag(n),matrix(1,1,t))
> A2 <- kronecker(FLA.constr,diag(t))
> delta.constr <- rbind(A1,A2)
```

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

```
> formula <- O ~ f(ID.FLA, model="generic1", Cmatrix=R.Leroux.block, constr=TRUE,
+                    hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif)),
+                    extraconstr=list(A=FLA.constr, e=rep(0,m))) +
+                  f(ID.SLA, model="generic1", Cmatrix=R.Leroux.SLA, constr=TRUE,
+                    hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
+                  f(ID.year, model="rw1", hyper=list(prec=list(prior=sdunif))) +
+                  f(ID.FLA.year, model="generic0", Cmatrix=R, rankdef=r.def,
+                    constr=TRUE, hyper=list(prec=list(prior=sdunif)),
+                    extraconstr=list(A=delta.constr, e=rep(0,n+m*t)))
```

where `N1,...,Nm` denotes the number of FLAs within the jth SLA, for $j = 1, \ldots, m$.

Two-level model D

```
> FLA.constr <- bdiag(matrix(1,1,N1),...,matrix(1,1,Nm))
> R <- kronecker(R.psi,R.gammaRW1)
> r.def <- m+t-1
> A1 <- kronecker(diag(m),matrix(1,1,t))
> A2 <- kronecker(matrix(1,1,m),diag(t))
> delta.constr <- rbind(A1,A2)
```

```
> formula <- O ~ f(ID.FLA, model="generic1", Cmatrix=R.Leroux.block, constr=TRUE,
+                    hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif)),
+                    extraconstr=list(A=FLA.constr, e=rep(0,m))) +
+                  f(ID.SLA, model="generic1", Cmatrix=R.Leroux.SLA, constr=TRUE,
+                    hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
+                  f(ID.year, model="rw1", hyper=list(prec=list(prior=sdunif))) +
+                  f(ID.SLA.year, model="generic0", Cmatrix=R, rankdef=r.def,
+                    constr=TRUE, hyper=list(prec=list(prior=sdunif)),
+                    extraconstr=list(A=delta.constr, e=rep(0,m+t)))
```

where `R` is the structure matrix $\mathbf{R}_\psi \otimes \mathbf{R}_\gamma$ and the linear constraints over the interaction term for the *two-level model D* are expressed as

$$\sum_{t=1}^{T} \delta_{jt} = 0, \quad \text{for} \quad j = 1, \ldots, m \iff (\mathbf{I}_m \otimes \mathbf{1}_T^{'})\boldsymbol{\delta} = 0,$$
$$\sum_{j=1}^{m} \delta_{jt} = 0, \quad \text{for} \quad t = 1, \ldots, T \iff (\mathbf{1}_m^{'} \otimes \mathbf{I}_T)\boldsymbol{\delta} = 0.$$

Two-level models E/F

For fitting models E and F, m independent random effects must be specified to model the municipality-level (FLA) spatial effect, replacing the term `f(ID.FLA,...)` by

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

```
f(ID.S1, model="generic1", Cmatrix=R.Leroux.S1, constr=TRUE, ...) +
...
f(ID.Sm, model="generic1", Cmatrix=R.Leroux.Sm, constr=TRUE, ...) +
```

where `R.Leroux.S1, ..., R.Leroux.Sm` are the $\mathbf{R}_{\xi_j}$ neighborhood matrices of FLAs within jth SLA, for $j = 1, \ldots, m$. As stated in Table 3.2, the same linear constraints as models C and D have to be imposed over the random effects of two-level models E and F respectively.

Finally, we run the INLA algorithm with a call to the `inla()` function as

```
> inla(formula, family="poisson", data=Data, E=E,
>      control.predictor=list(compute=TRUE, cdf=c(log(1))),
>      control.compute=list(dic=TRUE, cpo=TRUE, waic=TRUE),
>      control.inla=list(strategy="simplified.laplace"))
```

In addition to the marginal posterior distribution of the linear predictor $\log r_{it}$, the posterior probabilities $P(\log r_{it} > 0|\mathbf{O})$ are also computed by including the `control.predictor=list(compute=TRUE, cdf=c(log(1)))` argument.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

4 B-spline models to specify space-time interactions in Bayesian disease mapping

4.1 Introduction

In recent years, models incorporating splines have been used to smooth mortality risks in space and time. A sensible approach consists in using CAR priors for spatial random effects and B-splines (MacNab and Dean, 2001) or P-splines (Etxeberria et al., 2014) for temporal smoothing. Three-dimensional P-splines have been also used in this setting (Ugarte et al., 2010b, 2012a; Etxeberria et al., 2015) but from an empirical Bayes approach. In fact, these P-spline models have been embedded within a generalized linear mixed model (GLMM) framework and model fitting and inference has been carried out using the penalized quasi-likelihood technique (Breslow and Clayton, 1993). Splines have been also used to smooth risks in spatio-temporal disease mapping from a fully Bayesian approach, but their use has been mainly limited to model temporal rather than spatial effects (MacNab, 2007; MacNab and Gustafson, 2007). Although these models are quite flexible, model fitting has commonly relied on Markov chain Monte Carlo methods (Gilks et al., 1996) which may present some inconveniences. Apart from large computing time, there can be convergence problems which are essentially undetectable from the output analysis alone (see Gelman and Shirley, 2011). In addition, implementation of McMC may not be easy for practitioners (Knorr-Held and Rue, 2002; Schmid and Held, 2004), and large Monte Carlo errors could appear when fitting complex models (Schrödle et al., 2011).

The main objective of this chapter is to propose different possibilities of modelling the space-time interaction using (penalized or un-penalized) B-splines in Bayesian disease mapping as well as to specify the necessary constraints to make all the models identifiable. Specifically, we will consider one-dimensional, two-dimensional, and

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

three-dimensional B-splines. If the interest relies on analyzing the temporal evolution of risks in certain small areas (provinces, counties, municipalities or census tracts), fitting a temporal B-spline for each area could be preferable, making the splines of the areas spatially correlated if needed. On the other hand, if the focus is on studying the temporal evolution of the geographical pattern of mortality, a spatial surface for each time point using two-dimensional B or P-splines that can be (or not) temporally correlated could be fitted. Three-dimensional B-splines could be also used for both purposes (see Goicoa et al., 2016 for a discussion). Depending on the priors given to the spline coefficients for the spatial and temporal dimensions, the B-spline models can be classified according to different types of spatio-temporal interactions resembling the four types of interactions described by Knorr-Held (2000). As the models are in general not identifiable, constraints have to be defined. A spectral decomposition of the precision matrices of the B-splines coefficients is used to shed light on which and how many constraints are necessary (see for example Reich et al., 2006). The restrictions will be clearly established for the complete set of models. Because computational time could be an issue, all the proposed models, which are in the class of Gaussian Markov random fields, will be fitted using INLA. As the number of constraints needed to identify the models increases computing time, a look into these models becomes essential to define a fitting strategy.

The rest of the chapter is organized as follows. Section 4.2 describes the models proposed in this work. In Section 4.3 two dataset are used to illustrate the models' performance in terms of several model selection criteria and computing time. First, breast cancer mortality data in continental Spain is analyzed in Section 4.3.1, and secondly, an artificial dataset with a much higher number of small areas is considered in Section 4.3.2. The chapter finishes with a discussion.

4.2 B-spline models for spatio-temporal count data

In the following, different B-spline models are proposed for the log-relative risks using a fully Bayesian approach. The models are classified into three main groups depending on the dimension of the B-spline bases used to approximate the spatio-temporal effects.

4.2.1 One-dimensional splines for the space-time interaction

The model is defined as follows

$$\log r_{it} = \eta + \xi_i + f(x_t) + f_i(x_t) \quad \text{for} \quad \begin{matrix} i = 1, \ldots, n. \\ t = 1, \ldots, T. \end{matrix} \tag{4.1}$$

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

where η is an intercept, ξ_i is a spatially structured random effect, $f(x_t)$ is a temporal smooth function common to all areas, and $f_i(x_t)$ is an area specific temporal smooth function. As in the rest of this dissertation, a LCAR prior is considered for the spatial effects $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)'$, taking account of both spatially structured and unstructured variability (see Equation (1.2)).

The common temporal trend is specified as $f(\mathbf{x}_t) = \mathbf{B}_t\boldsymbol{\theta}^{(t)}$, where $\mathbf{B}_t$ is the temporal B-spline basis of dimension $T \times k_t$ (with k_t depending on the number of knots and the degree of the polynomials in the basis) obtained from the time covariate $\mathbf{x}_t = (x_1, \ldots, x_T)'$. To achieve smoothness, random walk priors on the unknown coefficients $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \ldots, \theta_{k_t}^{(t)})'$ are used (see Lang and Brezger, 2004), so that we assume

$$\boldsymbol{\theta}^{(t)} \sim N(\mathbf{0}, [\lambda_t \mathbf{P}_{k_t}]^-); \quad \mathbf{P}_{k_t} = \Delta_{d_t}^{'}\Delta_{d_t},$$

where Δ_{d_t} is the $k_t \times k_t$ difference matrix of order $d_t = 1$ or 2, the symbol $^-$ denotes the Moore-Penrose generalized inverse of a matrix, and λ_t represents the precision parameter. This is the Bayesian counterpart to placing penalties in the coefficients through finite order difference matrices in a frequentist approach (Eilers and Marx, 1996).

Finally, the area specific temporal smooth function is defined as $f_i(\mathbf{x}_t) = \mathbf{B}_{st}\boldsymbol{\theta}^{(st)}$, where $\mathbf{B}_{st} = \mathbf{I}_n \otimes \mathbf{B}_t$ is a block-diagonal matrix of dimension $nT \times nk_t$. Spatial and temporal correlation can be included using spatial CAR and temporal random walk priors on the coefficients $\boldsymbol{\theta}^{(st)} = (\theta_{11}^{(st)}, \ldots, \theta_{1k_t}^{(st)}, \ldots, \theta_{n1}^{(st)}, \ldots, \theta_{nk_t}^{(st)})'$ of the temporal B-spline. That is

$$\begin{aligned}
\boldsymbol{\theta}_{\cdot k}^{(st)} &= (\theta_{1k}^{(st)}, \ldots, \theta_{nk}^{(st)})' \sim N(\mathbf{0}, [\tau_{st}\mathbf{R}_\xi]^-); \quad \text{for} \quad k = 1, \ldots, k_t, \\
\boldsymbol{\theta}_{i\cdot}^{(st)} &= (\theta_{i1}^{(st)}, \ldots, \theta_{ik_t}^{(st)})' \sim N(\mathbf{0}, [\tau_{st}\mathbf{P}_{k_t}]^-); \quad \text{for} \quad i = 1, \ldots, n,
\end{aligned}$$

where $\mathbf{R}_\xi$ is the previously defined $n \times n$ spatial neighborhood matrix. Consequently $\boldsymbol{\theta}^{(st)} \sim N(\mathbf{0}, [\tau_{st}(\mathbf{R}_\xi \otimes \mathbf{P}_{k_t})]^-)$. Using these priors, the temporal B-splines become temporal P-splines and trends for neighbouring regions tend to be similar. If spatial correlation on the regression coefficients is ignored, then the area specific temporal trend will vary randomly. Depending on the priors for neighbouring coefficients in space and in time, different models arise resembling the four types of interaction models defined in Knorr-Held (2000). The structure matrices for the different type of interactions, and the priors for the coefficients in space and time are summarized in Table 4.1. Type I and Type III interactions correspond to spatially unstructured and spatially structured temporal B-splines respectively whereas Type II and Type IV interactions define spatially unstructured and spatially structured temporal P-splines respectively.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Table 4.1: Specification of the different types of structure matrices and the prior distribution over the regression coefficients of the interaction term $f_i(\mathbf{x}_t)$.

Interaction	Structure matrix	Model	Spatial prior	Temporal prior
Type I	$\mathbf{I}_n \otimes \mathbf{I}_{k_t}$	Temporal B-splines spatially unstructured	$\boldsymbol{\theta}_{\cdot k}^{(st)} \sim N(\mathbf{0}, \tau_{st}^{-1}\mathbf{I}_n)$	$\boldsymbol{\theta}_{i\cdot}^{(st)} \sim N(\mathbf{0}, \tau_{st}^{-1}\mathbf{I}_{k_t})$
Type II	$\mathbf{I}_n \otimes \mathbf{P}_{k_t}$	Temporal P-splines spatially unstructured	$\boldsymbol{\theta}_{\cdot k}^{(st)} \sim N(\mathbf{0}, \tau_{st}^{-1}\mathbf{I}_n)$	$\boldsymbol{\theta}_{i\cdot}^{(st)} \sim N(\mathbf{0}, [\tau_{st}\mathbf{P}_{k_t}]^-)$
Type III	$\mathbf{R}_\xi \otimes \mathbf{I}_{k_t}$	Temporal B-splines spatially structured	$\boldsymbol{\theta}_{\cdot k}^{(st)} \sim N(\mathbf{0}, [\tau_{st}\mathbf{R}_\xi]^-)$	$\boldsymbol{\theta}_{i\cdot}^{(st)} \sim N(\mathbf{0}, \tau_{st}^{-1}\mathbf{I}_{k_t})$
Type IV	$\mathbf{R}_\xi \otimes \mathbf{P}_{k_t}$	Temporal P-splines spatially structured	$\boldsymbol{\theta}_{\cdot k}^{(st)} \sim N(\mathbf{0}, [\tau_{st}\mathbf{R}_\xi]^-)$	$\boldsymbol{\theta}_{i\cdot}^{(st)} \sim N(\mathbf{0}, [\tau_{st}\mathbf{P}_{k_t}]^-)$

In the spatio-temporal model of Equation (4.1), the following sum-to-zero constraints are imposed to ensure identifiability between the main spatial and temporal effects and the model intercept

$$\sum_{i=1}^{n} \xi_i = 0 \quad \text{and} \quad \sum_{k=1}^{k_t} \theta_k^{(t)} = 0.$$

The identifiability issue arises because both the main spatial and the main temporal effects have implicitly defined an intercept which cannot be distinguished from the one included in the model. In addition, to avoid identifiability problems between the interaction term and the main effects, additional constraints must be imposed over the regression coefficients, because part of the interaction term is confounded with the main effects. The required identifiability constraints for the different types of space-time interactions are summarized in Table 4.2. Full details about how these constraints are derived are provided in Appendix 4A for the completely structured (Type IV) interaction model. The other cases are similarly derived.

4.2.2 Two-dimensional splines for the space-time interaction

Models incorporating two-dimensional smooth surfaces may be defined as

$$\log r_{it} = \eta + f(x_{1i}, x_{2i}) + \gamma_t + f_t(x_{1i}, x_{2i}) \quad \text{for} \quad \begin{matrix} i = 1, \ldots, n, \\ t = 1, \ldots, T, \end{matrix} \tag{4.2}$$

where η is an intercept, $f(x_{1i}, x_{2i})$ is a spatial smooth surface constant along the time periods, γ_t is a temporally structured random effect, and $f_t(x_{1i}, x_{2i})$ is a time specific spatial smooth surface. Here, first or second order random walk priors are

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Table 4.2: Identifiability constraints to fit one-dimensional temporal B-splines or P-splines described in Equation (4.1).

Interaction	$\Delta_{d_t} \equiv 1_{st}$ order penalty	$\Delta_{d_t} \equiv 2_{nd}$ order penalty
Type I	$\sum_{i=1}^{n} \xi_i = 0, \quad \sum_{k=1}^{k_t} \theta_k^{(t)} = 0$ $\sum_{i=1}^{n} \sum_{t=1}^{T} f_i(x_t) = 0$	$\sum_{i=1}^{n} \xi_i = 0, \quad \sum_{k=1}^{k_t} \theta_k^{(t)} = 0$ $\sum_{i=1}^{n} \sum_{t=1}^{T} f_i(x_t) = \sum_{i=1}^{n} \sum_{t=1}^{T} t\, f_i(x_t) = 0$
Type II	$\sum_{i=1}^{n} \xi_i = 0, \quad \sum_{k=1}^{k_t} \theta_k^{(t)} = 0$ $\sum_{k=1}^{k_t} \theta_{ik}^{(st)} = 0$, for $i = 1, \ldots, n$	$\sum_{i=1}^{n} \xi_i = 0, \quad \sum_{k=1}^{k_t} \theta_k^{(t)} = \sum_{k=1}^{k_t} k\, \theta_k^{(t)} = 0$ $\sum_{k=1}^{k_t} \theta_{ik}^{(st)} = 0$, for $i = 1, \ldots, n$
Type III	$\sum_{i=1}^{n} \xi_i = 0, \quad \sum_{k=1}^{k_t} \theta_k^{(t)} = 0$ $\sum_{i=1}^{n} \theta_{ik}^{(st)} = 0$, for $k = 1, \ldots, k_t$	$\sum_{i=1}^{n} \xi_i = 0, \quad \sum_{k=1}^{k_t} \theta_k^{(t)} = 0$ $\sum_{i=1}^{n} \theta_{ik}^{(st)} = 0$, for $k = 1, \ldots, k_t$
Type IV	$\sum_{i=1}^{n} \xi_i = 0, \quad \sum_{k=1}^{k_t} \theta_k^{(t)} = 0$ $\sum_{k=1}^{k_t} \theta_{ik}^{(st)} = 0$, for $i = 1, \ldots, n$ $\sum_{i=1}^{n} \theta_{ik}^{(st)} = 0$, for $k = 1, \ldots, k_t$	$\sum_{i=1}^{n} \xi_i = 0, \quad \sum_{k=1}^{k_t} \theta_k^{(t)} = 0$ $\sum_{k=1}^{k_t} \theta_{ik}^{(st)} = 0$, for $i = 1, \ldots, n$ $\sum_{i=1}^{n} \theta_{ik}^{(st)} = 0$, for $k = 1, \ldots, k_t$

considered for the temporal effects $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_T)'$, defined as

$$\boldsymbol{\gamma} \sim N(\mathbf{0}, [\tau_\gamma \mathbf{R}_\gamma]^-),$$

where $\mathbf{R}_\gamma = \Delta_{d_t}' \Delta_{d_t}$ is the $T \times T$ structure matrix of a RW1/RW2 and Δ_{d_t} is the difference matrix of order $d_t = 1$ or 2.

The constant spatial smooth surface is specified as $f(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{B}_s \boldsymbol{\theta}^{(s)}$, where $\mathbf{B}_s = \mathbf{B}_2 \square \mathbf{B}_1$ is the two-dimensional B-spline basis of dimension $n \times k_1 k_2$ (with k_1 and k_2 depending on the number of knots and the degree of the polynomials in the marginal B-spline bases $\mathbf{B}_1$ and $\mathbf{B}_2$), obtained from the row-wise Kronecker product (see Equation (2.2)) of the marginal bases for longitude $\mathbf{x}_1 = (x_{11}, \ldots, x_{1n})'$ and

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

latitude $\mathbf{x}_2 = (x_{21}, \ldots, x_{2n})'$. To achieve smoothness, different prior distributions can be placed over the regression coefficients $\boldsymbol{\theta}^{(s)} = (\theta_{11}^{(s)}, \ldots, \theta_{k_1k_2}^{(s)})'$, such that

$$\boldsymbol{\theta}^{(s)} \sim N(\mathbf{0}, \mathbf{P}_s^-),$$

where $\mathbf{P}_s$ is a precision matrix. If the smoothing is different in longitude and latitude, then the following precision matrix accounting for anisotropy can be used

$$\mathbf{P}_s = \lambda_{s_1}(\mathbf{I}_{k_2} \otimes \Delta_{d_1}'\Delta_{d_1}) + \lambda_{s_2}(\Delta_{d_2}'\Delta_{d_2} \otimes \mathbf{I}_{k_1}), \tag{4.3}$$

where $\mathbf{I}_{k_i}$ are identity matrices and Δ_{d_i} are difference matrices (order 1 or 2) of dimension $k_i \times k_i$, $i = 1, 2$. The hyperparameters λ_{s_1} and λ_{s_2} control the amount of smoothing in each direction. If the same amount of smoothing is expected in both directions, then the following isotropic precision matrix can be used instead

$$\mathbf{P}_s = \lambda_s(\mathbf{I}_{k_2} \otimes \Delta_{d_1}'\Delta_{d_1} + \Delta_{d_2}'\Delta_{d_2} \otimes \mathbf{I}_{k_1}). \tag{4.4}$$

This isotropic penalty works well if the predictor variables, longitude and latitude, are distributed reasonably evenly in a rectangular grid (Ruppert et al., 2003), as it is in our case study. However, even though the marginal B-spline bases and the tensor product bases are scaling invariant, the two-dimensional smoother is not scaling invariant with this penalty. This means that the relative importance of smoothness with respect to the predictors is arbitrary and depends on the scale. For example, if one variable is measured in a scale $[0, 20]$ and the other in a scale $[0, 5]$, the former will be penalized more heavily, something completely arbitrary. If both variables are transformed to the unit interval $[0, 1]$, both will be equally penalized, which again is arbitrary. This arbitrariness can lead to poor results. The scale invariance of the smoother is achieved using a different amount of smoothing in each direction letting the data estimate the smoothing parameters corresponding to each variable. See Wood (2006) and Wood et al. (2013) for an in depth discussion about scale invariance of tensor product smooths. Other types of precision (penalty) matrices can be also defined for two-dimensional P-splines. See Belitz and Lang (2008) for a detailed description of different precision or penalty matrices in a Bayesian setting.

Finally, a time specific spatial smooth surface is defined as

$$f_t(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{B}_{st}\boldsymbol{\theta}^{(st)},$$

where $\mathbf{B}_{st} = \mathbf{I}_T \otimes \mathbf{B}_s$ is a block-diagonal matrix of dimension $nT \times (k_1k_2)T$. Temporal and spatial correlations can be included putting random walk and spatial CAR priors on the coefficients $\boldsymbol{\theta}^{(st)} = (\theta_{11,1}^{(st)}, \ldots, \theta_{k_1k_2,1}^{(st)}, \ldots, \theta_{11,T}^{(st)}, \ldots, \theta_{k_1k_2,T}^{(st)})'$. That is, arranging the coefficients of the spatial bases by rows and columns (longitude and latitude)

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Table 4.3: Specification of the different types of structure matrices and the prior distribution over the regression coefficients of the interaction term $f_t(\mathbf{x}_1, \mathbf{x}_2)$.

Interaction	Structure matrix	Model	Temporal prior	Spatial prior
Type I	$\mathbf{I}_T \otimes \mathbf{I}_{k_1 k_2}$	Spatial B-splines temporally unstructured	$\boldsymbol{\theta}^{(st)}_{ij,\cdot} \sim N(\mathbf{0}, \tau_{st}^{-1}\mathbf{I}_T)$	$\boldsymbol{\theta}^{(st)}_{\cdot\cdot,t} \sim N(\mathbf{0}, \tau_{st}^{-1}\mathbf{I}_{k_1 k_2})$
Type II	$\mathbf{R}_\gamma \otimes \mathbf{I}_{k_1 k_2}$	Spatial B-splines temporally structured	$\boldsymbol{\theta}^{(st)}_{ij,\cdot} \sim N(\mathbf{0}, [\tau_{st}\mathbf{R}_\gamma]^-)$	$\boldsymbol{\theta}^{(st)}_{\cdot\cdot,t} \sim N(\mathbf{0}, \tau_{st}^{-1}\mathbf{I}_{k_1 k_2})$
Type III	$\mathbf{I}_T \otimes \mathbf{P}_s$	Spatial P-splines temporally unstructured	$\boldsymbol{\theta}^{(st)}_{ij,\cdot} \sim N(\mathbf{0}, \tau_{st}^{-1}\mathbf{I}_T)$	$\boldsymbol{\theta}^{(st)}_{\cdot\cdot,t} \sim N(\mathbf{0}, [\tau_{st}\mathbf{P}_s]^-)$
Type IV	$\mathbf{R}_\gamma \otimes \mathbf{P}_s$	Spatial P-splines temporally structured	$\boldsymbol{\theta}^{(st)}_{ij,\cdot} \sim N(\mathbf{0}, [\tau_{st}\mathbf{R}_\gamma]^-)$	$\boldsymbol{\theta}^{(st)}_{\cdot\cdot,t} \sim N(\mathbf{0}, [\tau_{st}\mathbf{P}_s]^-)$

for each time point $t = 1, \ldots, T$

$$\begin{aligned} \boldsymbol{\theta}^{(st)}_{ij,\cdot} &= (\theta^{(st)}_{ij,1}, \ldots, \theta^{(st)}_{ij,T})' \sim N(\mathbf{0}, [\tau_{st}\mathbf{R}_\gamma]^-); \quad \text{for} \quad \begin{matrix} i = 1, \ldots, k_1, \\ j = 1, \ldots, k_2, \end{matrix} \\ \boldsymbol{\theta}^{(st)}_{\cdot\cdot,t} &= (\theta^{(st)}_{11,t}, \ldots, \theta^{(st)}_{k_1 k_2,t})' \sim N(\mathbf{0}, [\tau_{st}\mathbf{P}_s]^-); \quad \text{for} \quad t = 1, \ldots, T. \end{aligned}$$

Consequently, $\boldsymbol{\theta}^{(st)} \sim N(\mathbf{0}, [\tau_{st}(\mathbf{R}_\gamma \otimes \mathbf{P}_s)]^-)$. Using these priors, the two-dimensional spatial B-splines become two-dimensional spatial P-splines and spatial surfaces from consecutive years tend to be similar. Analogously to the models presented in the previous section, if temporal correlation on the regression coefficients is ignored, then independent two-dimensional risk surfaces are considered for each year. The structure matrices for the different types of interactions, and the priors for the coefficients in space and time are summarized in Table 4.3. Note that Type I and Type II interactions correspond to temporally unstructured or temporally structured spatial B-splines, whereas Type III and Type IV interactions define temporally unstructured or temporally structured spatial P-splines respectively. In these cases, both isotropic and anisotropic precision matrices can be considered.

In these models, an intercept is implicitly included in the spatial surface and the main temporal effect, consequently identifiability constraints must be also imposed to ensure identifiability between the main spatial and temporal effects and the model intercept. The following sum-to-zero constraints have to be specified

$$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2} \theta^{(s)}_{ij} = 0 \quad \text{and} \quad \sum_{t=1}^{T} \gamma_t = 0.$$

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Table 4.4: Identifiability constraints to fit two-dimensional B-splines or P-splines described in Equation (4.2). Here first order penalties for longitude and latitude are considered for P-splines.

	$\Delta_{d_1}, \Delta_{d_2} \equiv 1_{st}$ order penalty	
Interaction	$\Delta_{d_t} \equiv 1_{st}$ order penalty	$\Delta_{d_t} \equiv 2_{nd}$ order penalty
Type I	$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)} = 0, \quad \sum_{t=1}^{T}\gamma_t = 0$ $\sum_{i=1}^{n}\sum_{t=1}^{T} f_t(x_{1i}, x_{2i}) = 0$	$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)} = 0, \quad \sum_{t=1}^{T}\gamma_t = 0$ $\sum_{i=1}^{n}\sum_{t=1}^{T} f_t(x_{1i}, x_{2i}) = \sum_{i=1}^{n}\sum_{t=1}^{T} t\, f_t(x_{1i}, x_{2i}) = 0$
Type II	$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)} = 0, \quad \sum_{t=1}^{T}\gamma_t = 0$ $\sum_{t=1}^{T}\theta_{ij,t}^{(st)} = 0$, for $i = 1,\ldots,k_1$, $j = 1,\ldots,k_2$	$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)} = 0, \quad \sum_{t=1}^{T}\gamma_t = \sum_{t=1}^{T} t\,\gamma_t = 0$ $\sum_{t=1}^{T}\theta_{ij,t}^{(st)} = 0$, for $i = 1,\ldots,k_1$, $j = 1,\ldots,k_2$
Type III	$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)} = 0, \quad \sum_{t=1}^{T}\gamma_t = 0$ $\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij,t}^{(st)} = 0$, for $t = 1,\ldots,T$	$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)} = 0, \quad \sum_{t=1}^{T}\gamma_t = 0$ $\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij,t}^{(st)} = 0$, for $t = 1,\ldots,T$
Type IV	$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)} = 0, \quad \sum_{t=1}^{T}\gamma_t = 0$ $\sum_{t=1}^{T}\theta_{ij,t}^{(st)} = 0$, for $i = 1,\ldots,k_1$, $j = 1,\ldots,k_2$ $\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij,t}^{(st)} = 0$, for $t = 1,\ldots,T$	$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)} = 0, \quad \sum_{t=1}^{T}\gamma_t = 0$ $\sum_{t=1}^{T}\theta_{ij,t}^{(st)} = 0$, for $i = 1,\ldots,k_1$, $j = 1,\ldots,k_2$ $\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij,t}^{(st)} = 0$, for $t = 1,\ldots,T$

In addition, to avoid identifiability problems between the interaction term and the main effects, new constraints must be imposed over the regression coefficients, because part of the interaction term repeats the main effects. The required identifiability constraints are summarized in Table 4.4 (differences of first order in the spatial precision matrix) and Table 4.5 (differences of second order in the spatial precision matrix) for the four types of space-time interactions. Again, full details about how these constraints are derived are provided in Appendix 4A for the completely structured (Type IV) interaction model. The other cases are similarly derived.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Table 4.5: Identifiability constraints to fit two-dimensional B-splines or P-splines described in Equation (4.2). Here second order penalties for longitude and latitude are considered for P-splines.

	$\Delta_{d_1}, \Delta_{d_2} \equiv 2_{nd}$ order penalty	
Interaction	$\Delta_{d_t} \equiv 1_{st}$ order penalty	$\Delta_{d_t} \equiv 2_{nd}$ order penalty
Type I	$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)} = 0, \quad \sum_{t=1}^{T}\gamma_t = 0$ $\sum_{i=1}^{n}\sum_{t=1}^{T} f_t(x_{1i}, x_{2i}) = 0$ $\sum_{i=1}^{n}\sum_{t=1}^{T} x_{1i}\, f_t(x_{1i}, x_{2i}) = \sum_{i=1}^{n}\sum_{t=1}^{T} x_{2i}\, f_t(x_{1i}, x_{2i}) = 0$ $\sum_{i=1}^{n}\sum_{t=1}^{T} x_{1i}\, x_{2i}\, f_t(x_{1i}, x_{2i}) = 0$	$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)} = 0, \quad \sum_{t=1}^{T}\gamma_t = 0$ $\sum_{i=1}^{n}\sum_{t=1}^{T} f_t(x_{1i}, x_{2i}) = \sum_{i=1}^{n}\sum_{t=1}^{T} t\, f_t(x_{1i}, x_{2i}) = 0$ $\sum_{i=1}^{n}\sum_{t=1}^{T} x_{1i}\, f_t(x_{1i}, x_{2i}) = \sum_{i=1}^{n}\sum_{t=1}^{T} x_{2i}\, f_t(x_{1i}, x_{2i}) = 0$ $\sum_{i=1}^{n}\sum_{t=1}^{T} x_{1i}\, x_{2i}\, f_t(x_{1i}, x_{2i}) = 0$
Type II	$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)} = 0, \quad \sum_{t=1}^{T}\gamma_t = 0$ $\sum_{t=1}^{T}\theta_{ij,t}^{(st)} = 0$, for $\begin{matrix} i = 1,\dots,k_1 \\ j = 1,\dots,k_2 \end{matrix}$	$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)} = 0, \quad \sum_{t=1}^{T}\gamma_t = \sum_{t=1}^{T} t\,\gamma_t = 0$ $\sum_{t=1}^{T}\theta_{ij,t}^{(st)} = 0$, for $\begin{matrix} i = 1,\dots,k_1 \\ j = 1,\dots,k_2 \end{matrix}$
Type III	$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)} = \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} i\,\theta_{ij}^{(s)} = 0$ $\sum_{i=1}^{k_1}\sum_{j=1}^{k_2} j\,\theta_{ij}^{(s)} = \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} i\,j\,\theta_{ij}^{(s)} = 0$ $\sum_{t=1}^{T}\gamma_t = 0$ $\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij,t}^{(st)} = 0$, for $t = 1,\dots,T$	$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)} = \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} i\,\theta_{ij}^{(s)} = 0$ $\sum_{i=1}^{k_1}\sum_{j=1}^{k_2} j\,\theta_{ij}^{(s)} = \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} i\,j\,\theta_{ij}^{(s)} = 0$ $\sum_{t=1}^{T}\gamma_t = 0$ $\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij,t}^{(st)} = 0$, for $t = 1,\dots,T$
Type IV	$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)} = 0, \quad \sum_{t=1}^{T}\gamma_t = 0$ $\sum_{t=1}^{T}\theta_{ij,t}^{(st)} = 0$, for $\begin{matrix} i = 1,\dots,k_1 \\ j = 1,\dots,k_2 \end{matrix}$ $\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij,t}^{(st)} = 0$, for $t = 1,\dots,T$	$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)} = 0, \quad \sum_{t=1}^{T}\gamma_t = 0$ $\sum_{t=1}^{T}\theta_{ij,t}^{(st)} = 0$, for $\begin{matrix} i = 1,\dots,k_1 \\ j = 1,\dots,k_2 \end{matrix}$ $\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij,t}^{(st)} = 0$, for $t = 1,\dots,T$

4.2.3 Three-dimensional P-splines

The models described above avoid the use of three-dimensional B-spline bases, achieving smoothness over the regression coefficients by placing different types of temporally or spatially structured priors on the coefficients. We want to determine if these models are good alternatives to three-dimensional P-spline models, which

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

are briefly described below in a Bayesian setting in contrast to Ugarte et al. (2010b) and Ugarte et al. (2012a) that used these models in an empirical Bayes disease mapping setting.

Interaction P-spline model

To achieve smoothness, the following prior distribution is considered for the regression coefficients $\boldsymbol{\theta}^{(st)}$ of the Interaction P-spline model defined in Equation (2.1)

$$\boldsymbol{\theta}^{(st)} \sim N(\mathbf{0}, \mathbf{P}_{st}^{-}),$$

where the precision matrix $\mathbf{P}_{st}$ is the penalty matrix used by Ugarte et al. (2010b) in an empirical Bayes setting (see Section 2.2.1). In a fully Bayesian approach, the hyperparameters τ_1, τ_2 and τ_t control the amount of smoothing in each direction. Note that three different smoothing parameters are required to get scaling invariant smooths, as time and space are covariates measured in different units and any transformation of the covariates to the same scale is arbitrary. In this model there is a single identifiability issue related to the intercept. Hence, only the $\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t}\theta_{ijk} = 0$ constraint need to be imposed.

ANOVA-type P-spline model

To achieve smoothness, the following prior distributions are considered for the regression coefficients $\boldsymbol{\theta}^{(s)}$, $\boldsymbol{\theta}^{(t)}$ and $\boldsymbol{\theta}^{(st)}$ of the ANOVA-type P-spline model defined in Equation (2.3)

$$\begin{aligned}
\boldsymbol{\theta}^{(s)} &\sim N(\mathbf{0}, [\lambda_{s_1}(\mathbf{I}_{k_2} \otimes \Delta_{d_1}'\Delta_{d_1}) + \lambda_{s_2}(\Delta_{d_2}'\Delta_{d_2} \otimes \mathbf{I}_{k_1})]^{-}),\\
\boldsymbol{\theta}^{(t)} &\sim N(\mathbf{0}, [\lambda_t \mathbf{P}_{k_t}]^{-}),\\
\boldsymbol{\theta}^{(st)} &\sim N(\mathbf{0}, \mathbf{P}_{st}^{-}).
\end{aligned}$$

Identifiability constraints using difference matrices of first or second order are summarized in Table 4.6. The number of constraints depends on the order of the penalty, and they are required because the interaction term includes linear terms which also enter in the main effects. See Appendix 4A for details about how these constraints are derived using a spectral decomposition of the precision matrices.

All the models presented in this chapter circumvent the identifiability problems if appropriate constraints are imposed over the spline regression coefficients. Other authors have also dealt with this problem when fitting Bayesian splines. For example, Lang and Brezger (2004) suggest imposing constraints over the unknown smooth functions, centering them in every iteration of the McMC algorithm. Very recently,

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Table 4.6: Identifiability constraints for the ANOVA-type P-spline model.

	$\Delta_{d_3} \equiv 1_{st}$ order penalty	$\Delta_{d_3} \equiv 2_{nd}$ order penalty
$\Delta_{d_1}, \Delta_{d_2} \equiv 1_{st}$ order penalty	$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)}=0, \quad \sum_{k=1}^{k_t}\theta_{k}^{(t)}=0$ $\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t}\theta_{ijk}^{(st)}=0$	$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)}=0, \quad \sum_{k=1}^{k_t}\theta_{k}^{(t)}=0$ $\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t}\theta_{ijk}^{(st)}=\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t}k\theta_{ijk}^{(st)}=0$
$\Delta_{d_1}, \Delta_{d_2} \equiv 2_{nd}$ order penalty	$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)}=0, \quad \sum_{k=1}^{k_t}\theta_{k}^{(t)}=0$ $\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t}\theta_{ijk}^{(st)}=\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t}i\theta_{ijk}^{(st)}=0$ $\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t}j\theta_{ijk}^{(st)}=\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t}ij\theta_{ijk}^{(st)}=0$	$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)}=0, \quad \sum_{k=1}^{k_t}\theta_{k}^{(t)}=0$ $\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t}\theta_{ijk}^{(st)}=\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t}i\theta_{ijk}^{(st)}=0$ $\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t}j\theta_{ijk}^{(st)}=\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t}ij\theta_{ijk}^{(st)}=0$ $\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t}k\theta_{ijk}^{(st)}=0$

similar sum-to-zero constraints are described for additive P-splines in Ventrucci and Rue (2016).

4.3 Illustration

Two dataset will be analyzed in this section using all the models described in this chapter. First, female breast cancer mortality data in continental Spain during the period 1990-2010 is used as illustration. Secondly, a simulated data for the municipalities of Navarre and Basque Country (Spain) has been considered, where the number of small areas is quite higher.

Improper uniform prior distributions were given to the square root inverse of the precision parameters, while a vague zero mean normal distribution with a precision close to zero (0.00001) was considered for the intercept (η). Finally, a Uniform(0,1) distribution has been used for the spatial smoothing parameter λ_ξ of the LCAR prior. The `R` code to fit some of these B-spline models using the `R-INLA` library is also included in Appendix 4B.

4.3.1 Breast cancer mortality data in continental Spain

Breast cancer mortality (ICD-10 code 50) represents the 17% of female cancer-related deaths in Spain during the period 1990-2010 and the 3% of overall female mortality in 2012, being also the leading cause of death by cancer in European females (Ferlay et al., 2013). Recent studies have shown that breast cancer in women

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

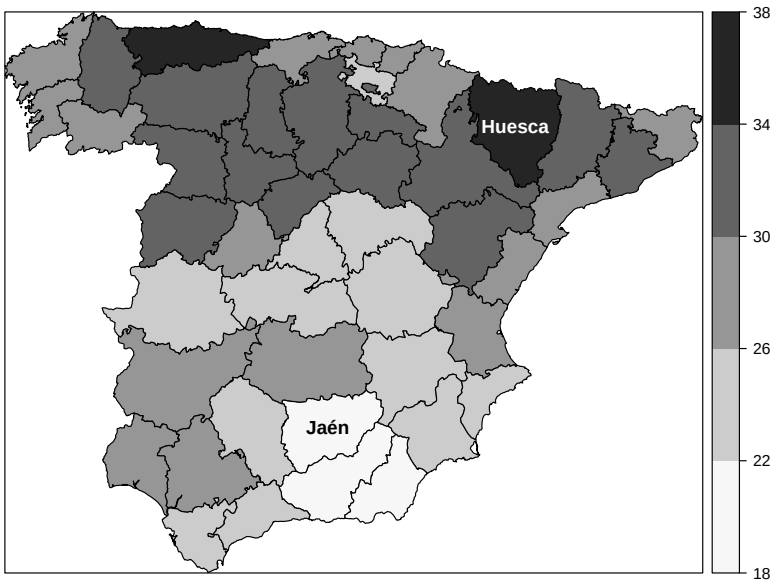


Figure 4.1: Breast cancer mortality rates in Spanish provinces (per 100000 female inhabitants). Period 1990-2010.

has registered a decline in mortality during the last years, although the temporal evolution does not seem to be the same for different regions of Spain (López-Abente et al., 2014). Therefore, it is crucial to perform studies where region-specific trends are analysed in addition to main spatial effects and temporal patterns common to all areas (Spanish provinces in this case).

A total of 114748 deaths were recorded by the Spanish National Epidemiology Center during the period 1990-2010 for the 47 provinces in continental Spain. The number of expected cases were computed using the Spanish population and standardizing by age. The number of expected deaths ranges from 16 to 930 while the number of observed deaths varies from 5 to 874, Huesca being the province with the highest crude mortality rate, and Jaén the province with the lowest one (36.8 and 20.1 per 100000 female inhabitants respectively). A map with breast cancer mortality rates during the whole period in Spanish provinces is displayed in Figure 4.1.

The models described in Section 4.2 have been fitted, combining different priors in space/time for the main effects as well as the four types of interactions for the spatio-temporal term. A total of 32 different models were considered. Specifically

- One dimensional temporal B-or-P splines spatially unstructured/structured (first and second order penalties) -8 models-. Six equidistant internal knots have been considered to construct the temporal cubic B-spline basis $\mathbf{B}_t$ of dimension $T \times k_t$ (21×8).

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

- Two-dimensional spatial B-or-P splines temporally unstructured/structured (first and second order penalties; isotropic and anisotropic models) -16 models-. Eight equidistant internal knots have been considered to construct the spatial cubic B-spline bases (longitude and latitude). The final spatial basis ($\mathbf{B}_s = \mathbf{B}_2 \square \mathbf{B}_1$) is of dimension $n \times k_1 k_2$ (47×100).

- Three dimensional P-splines (interaction and ANOVA-type with first and second order penalties) -8 models-. The same number of equidistant internal knots have been used to construct the temporal and spatial cubic B-spline bases (six and eight respectively).

Table 4.7 shows the results from fitting the 32 models using full Laplace approximations, including usual model selection criteria such as the Deviance Information Criterion (DIC), the corrected version of the DIC (DICc), the logarithmic score (LS) and the Watanabe-Akaike information criterion (WAIC). The estimation of the posterior marginals of the hyperparameters have been improved using the grid integration strategy (Martins et al., 2013) with the `inla.hyperpar` function. The models with area specific one-dimensional temporal P-splines perform better in this scenario in terms of model fitting and complexity. A model including a CAR spatial effect, a P-spline for the temporal effect (first order penalty), and a spatially correlated temporal P-spline for the interaction (Type IV interaction, $\mathbf{R}_\xi \otimes \mathbf{P}_{k_t}$, in Table 4.1) is selected as the best model. The ANOVA-type P-spline model with second order differences in the spatial precision matrix (for both longitude and latitude) and first order differences in the temporal precision matrix is the best model among those with three-dimensional B-spline bases. Regarding computing time, spatially unstructured or spatially structured temporal B or P-splines are the fastest models. Three main factors may increase computing time: the dimension of the random field, the number of variance/precision parameters, and the number of constraints. When a "small-to-moderate" number of small areas is analyzed (47 provinces), the length of the interaction random effect $\boldsymbol{\theta}^{(st)}$ is lower for area specific temporal P-spline models than for the rest of models, since $(n \times k_t) \ll (k_1 k_2 \times k_t) < (k_1 k_2 \times T)$. The number of variance/precision parameters is relatively small for the spatially correlated temporal P-splines. The number of constraints is in general smaller for the three-dimensional P-spline models.

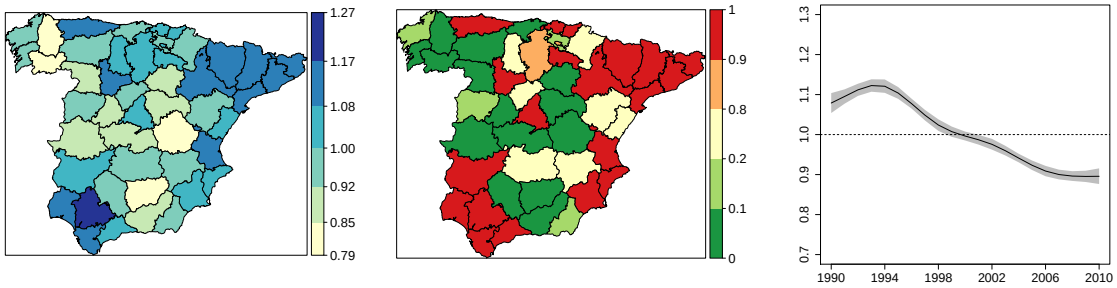
In the following, results for the best model are presented. Figure 4.2a shows the spatial mortality risk pattern $\zeta_i = \exp(\xi_i^*)$ (constant during the whole period) associated to each province, while Figure 4.2b displays the posterior probabilities that these risks are greater than one. The temporal risk pattern $\exp(\gamma_t^*)$ common to all provinces is visualized in Figure 4.2c, as well as the 95% credibility interval. Regarding Figure 4.2c, the decrease in breast cancer mortality is a generalized phenomenon documented in the literature (see Ugarte et al., 2010a). In Spain, the

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Table 4.7: Model selection criteria (DIC, DICc, WAIC and logarithmic score) and computational time (in seconds) from fitted models in the analysis of breast cancer mortality data in Spain. Full Laplace approximation.

	Spatially correlated one-dimensional temporal P-splines									
	$\mathbf{P}_k = \Delta_1'\Delta_1$					$\mathbf{P}_k = \Delta_2'\Delta_2$				
Interaction	DIC	DICc	WAIC	LS	time	DIC	DICc	WAIC	LS	time
Type I	7234.3	7256.1	7242.9	3624.4	23	7235.7	7257.8	7244.6	3625.3	22
Type II	7219.6	7235.3	7226.1	3615.0	39	7233.4	7252.3	7241.4	3623.3	39
Type III	7217.9	7233.3	7224.8	3614.3	20	7219.3	7234.9	7226.5	3615.2	20
Type IV	**7206.5**	**7218.7**	**7211.4**	**3607.1**	45	7228.3	7245.7	7235.4	3619.9	46
	Temporally correlated two-dimensional spatial P-splines									
Isotropic penalty	$\mathbf{P}_s = \lambda_s(\mathbf{I}_{k_2} \otimes \Delta_{d_1}'\Delta_{d_1} + \Delta_{d_2}'\Delta_{d_2} \otimes \mathbf{I}_{k_1})$									
	$\mathbf{R}_\gamma = \Delta_1'\Delta_1$					$\mathbf{R}_\gamma = \Delta_2'\Delta_2$				
Interaction	DIC	DICc	WAIC	LS	time	DIC	DICc	WAIC	LS	time
Type I	7303.3	7365.9	7314.1	3670.0	72	7301.4	7360.7	7312.0	3668.3	78
Type II	7238.1	7260.5	7244.5	3625.6	611	**7233.1**	**7247.8**	**7242.1**	**3624.1**	596
Type III	7318.2	7366.2	7335.9	3677.6	140	7316.4	7362.1	7333.9	3676.2	143
Type IV	7304.4	7360.5	7316.7	3669.0	1067	7275.0	7321.2	7285.9	3654.1	1075
Anisotropic penalty	$\mathbf{P}_s = \lambda_{s_1}(\mathbf{I}_{k_2} \otimes \Delta_{d_1}'\Delta_{d_1}) + \lambda_{s_2}(\Delta_{d_2}'\Delta_{d_2} \otimes \mathbf{I}_{k_1})$									
	$\mathbf{R}_\gamma = \Delta_1'\Delta_1$					$\mathbf{R}_\gamma = \Delta_2'\Delta_2$				
Interaction	DIC	DICc	WAIC	LS	time	DIC	DICc	WAIC	LS	time
Type I	7302.0	7365.5	7312.5	3669.2	148	7300.5	7361.7	7310.7	3667.8	185
Type II	7236.0	7258.3	7242.2	3624.4	1135	**7234.6**	**7249.8**	**7243.6**	**3624.6**	1095
Type III	7315.6	7367.3	7332.4	3676.4	279	7313.4	7362.4	7330.0	3674.7	279
Type IV	7297.4	7357.6	7307.4	3664.9	2414	7272.3	7321.0	7282.7	3652.8	2038
	Interaction P-spline model									
	$\Delta_{d_t} \equiv 1_{st}$ order penalty					$\Delta_{d_t} \equiv 2_{nd}$ order penalty				
	DIC	DICc	WAIC	LS	time	DIC	DICc	WAIC	LS	time
$\Delta_{d_1}, \Delta_{d_2} \equiv 1_{st}$ order penalty	7256.7	7271.6	7266.1	3635.0	71	7285.2	7298.9	7297.3	3650.6	67
$\Delta_{d_1}, \Delta_{d_2} \equiv 2_{nd}$ order penalty	**7251.6**	**7266.5**	**7260.4**	**3632.1**	75	7281.6	7295.5	7293.3	3648.7	70
	ANOVA-type P-spline model									
	$\Delta_{d_t} \equiv 1_{st}$ order penalty					$\Delta_{d_t} \equiv 2_{nd}$ order penalty				
	DIC	DICc	WAIC	LS	time	DIC	DICc	WAIC	LS	time
$\Delta_{d_1}, \Delta_{d_2} \equiv 1_{st}$ order penalty	7236.9	7246.9	7244.3	3623.3	208	7243.1	7251.9	7250.8	3626.6	202
$\Delta_{d_1}, \Delta_{d_2} \equiv 2_{nd}$ order penalty	**7231.4**	**7241.1**	**7238.3**	**3620.3**	240	7240.5	7249.5	7247.8	3625.1	229

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa



(a) Map of the spatial pattern of mortality risks $\zeta_i = \exp(\xi_i^*)$

(b) Map of posterior probabilities $P(\zeta_i > 1|\mathbf{O})$

(c) Common temporal trend of breast cancer mortality risk

Figure 4.2: Spatial and temporal patterns of female breast cancer mortality relative risks in Spanish provinces.

economic growth and social transformation of recent decades have allowed the improvement of treatments, advances in screening, and early diagnosis which in turn have led to a reduction in mortality from the beginning of the nineties onward.

Finally, Figure 4.3 shows the temporal evolution of female breast cancer mortality for six selected provinces ($\exp(\delta_{it}^*)$) and the corresponding 95% two-sided credible intervals. The colors used in this figure are those associated to the legend of the map shown in Figure 4.2b but for $P(\exp(\delta_{it}^*) > 1)$. The percentage of variability of the overall risk explained by the estimated spatial, temporal, and spatio-temporal patterns is about 56%, 37% and 7% respectively. For the sake of comparison, the percentages of explained variability for a two-dimensional anisotropic P-spline model (Type II interaction) and the ANOVA-type P-spline model (the models with better model selection criteria within their class of models) are 55% (spatial), 31% (temporal), and 14% (spatio-temporal), and 57% (spatial), 37% (temporal), and 6% (spatio-temporal) respectively.

In general, rather similar estimated relative risks are obtained. Figure 4.4 shows boxplots of the relative risk estimates (each boxplot within the three figures represents the relative risks of the different provinces in one year) for the spatially structured one-dimensional temporal P-spline model (a), the temporally structured two-dimensional (anisotropic) spatial B-spline model (a Type II interaction) (b), and the three-dimensional ANOVA-type P-spline model (c). Slight differences are observed on the underlying temporal pattern derived from the estimated risks, the ANOVA-type P-spline model producing the smoothest pattern.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

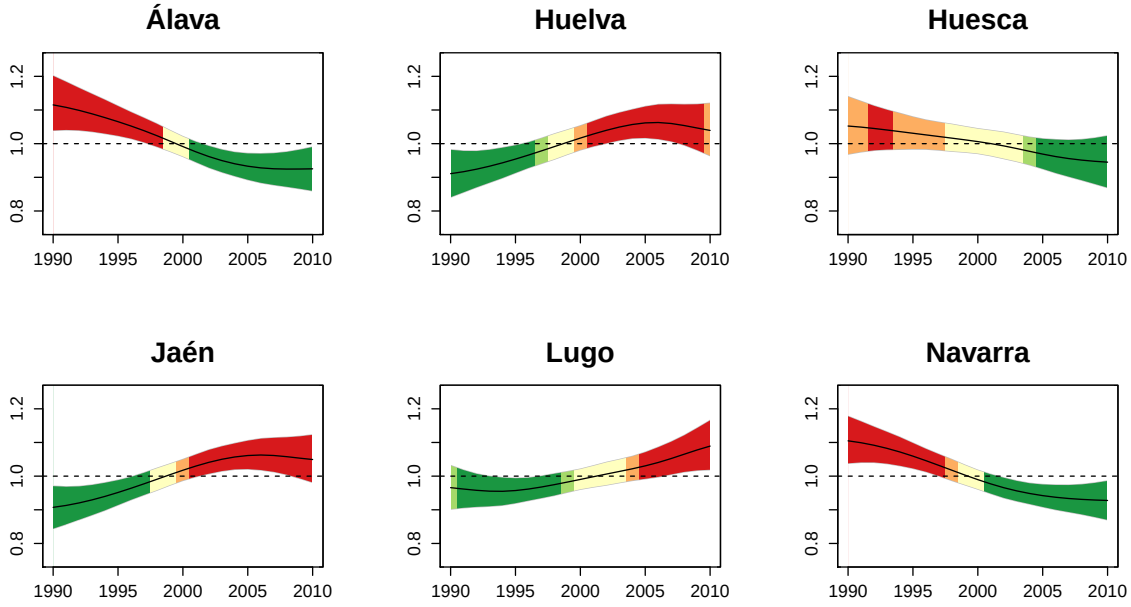


Figure 4.3: Temporal evolutions of six selected Spanish provinces, $\exp(\delta_{it}^*)$, and 95% two-sided credible intervals. The colors used in this figure are those associated to the legend of the map shown in Figure 4.2b but for $P(\exp(\delta_{it}^*) > 1)$.

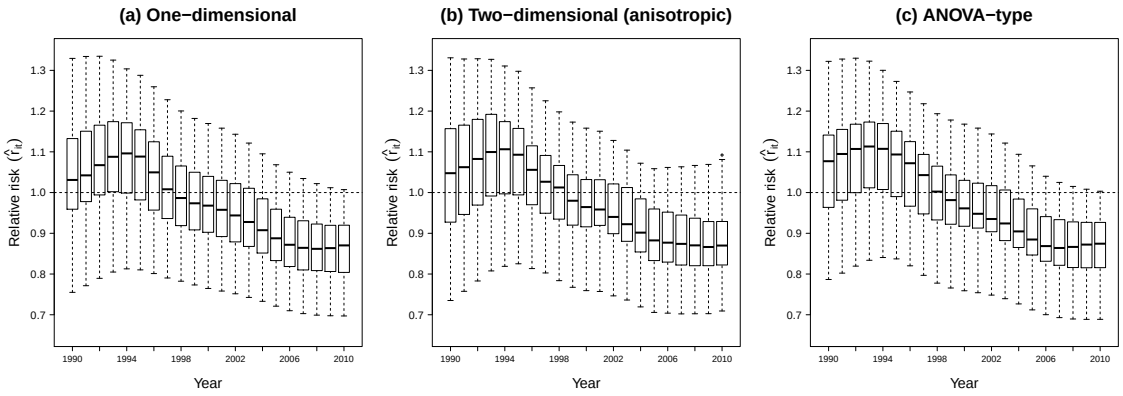


Figure 4.4: Boxplots of the relative risk estimates for the spatially structured one-dimensional temporal P-spline model (a), the temporally structured two-dimensional (anisotropic) spatial P-spline model with a Type II interaction (b), and the three-dimensional ANOVA-type P-spline model (c).

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

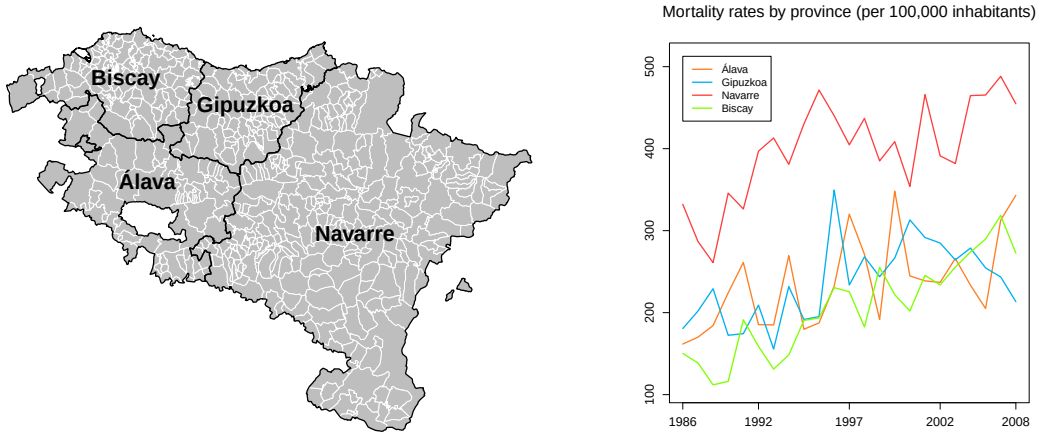


Figure 4.5: Map of the municipalities of Navarre and Basque Country (left) and temporal trends of simulated mortality rates per provinces (right)

4.3.2 Simulated data for the municipalities of Navarre and Basque Country

To further analyze the effect of the number of areas in the models described in Section 4.2, an artificial data set with $n = 501$ areas and $T = 23$ periods has been constructed based on the municipalities of four provinces in northern Spain (Álava, Gipuzcoa, Navarra, and Vizcaya; see left side of Figure 4.5). Brain cancer mortality data on these municipalities has been already analyzed in Section 3.3.1, where a second level of spatial aggregation was introduced (basic health areas) to smooth risks in space and time. This model shows better results than the usual spatio-temporal CAR models described in Knorr-Held (2000), due to the lack of municipality level space-time interaction in the data. To avoid this problem when fitting the Bayesian P-spline models, we decided to generate counts that maintains the spatial and temporal dependence structures obtained from the analysis of brain cancer data but including a stronger space-time interaction effect.

To do this, O^*_{it} counts have been generated from a Poisson distribution with mean $e^*_{it}r^*_{it}$ for the areas $i = 1, \ldots, 501$ and years $t = 1986, \ldots, 2008$. The following log-risk surface has been considered,

$$\log r^*_{it} = \xi^*_i + \gamma^*_t + \delta^*_{it}$$

where ξ^*_i and γ^*_t are the estimated spatial and temporal patterns from brain cancer mortality data analysis obtained in Section 3.3.1, while δ^*_{it} is a completely struc-

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

tured (Type IV) interaction term simulated from a multivariate normal distribution $\boldsymbol{\delta}^* \sim N(\mathbf{0}, [\tau_\delta(\mathbf{R}_\xi \otimes \mathbf{R}_\gamma)]^-)$. Here, $\mathbf{R}_\xi$ is the $n \times n$ spatial neighborhood matrix of the $n = 501$ municipalities and $\mathbf{R}_\gamma$ represents the $T \times T$ precision matrix of a RW1 for the $T = 23$ time periods. The precision component τ_δ has been properly chosen so that the interaction term explains around the 10% of the total variability. Finally, since brain cancer deaths represents around the 3% of the total cancer deaths in these provinces, the number of expected cases has been defined as $e^*_{it} = O_{it}/(0.03r^*_{it})$, where O_{it} represents the real brain cancer mortality counts. That is, the simulated counts will be similar to global cancer mortality data, assuming that the main spatial distribution and temporal evolution behave like in brain cancer data. Temporal trends of simulated mortality rates per province are shown in the right side of Figure 4.5.

To fit the data, the number of columns in the marginal spatial bases were $k_1 = k_2 = 13$ and the number of columns in the temporal basis was $k_t = 9$. After a preliminary analysis, first order difference matrices (Δ_{d_1} and Δ_{d_2}) were considered for the spatial precision matrix $\mathbf{P_s}$, while both RW1 and RW2 prior distributions have been considered for the temporal effect. The obtained results are summarized in Table 4.8.

As expected, the computational burden increases in all the models when the number of areas is big. In particular, the computational time to fit spatially correlated temporal P-splines for the space-time interaction increases substantially with respect to the fitting of temporally correlated spatial B-splines (11,475 minutes versus 514 minutes using a simplified Laplace approximation). The reason is that the number of random effects and basis coefficients as well as the number of constraints are higher for the former model, i.e, $(n + k_t + nk_t) > (k_1k_2 + T + k_1k_2T)$, and $(n + k_t) > (k_1k_2 + T)$. In this setting, the ANOVA-type P-spline model is very competitive as it takes 69 minutes to fit the data. Note that it requires seven constraints at most.

4.4 Discussion

In this chapter (penalized or unpenalized) B-splines are proposed to model space-time interactions in Bayesian disease mapping. The different models can be classified in three main groups according to the dimension of the B-spline bases used to approximate the spatio-temporal effects. Firstly, one-dimensional B-spline models have been defined, where specific temporal smooth functions are estimated for each small area. Secondly, two-dimensional B-spline models are studied, where specific spatial surfaces are estimated for each time period. Finally, three-dimensional P-spline models are described, where a smooth function of longitude, latitude, and time is defined to model the log-relative risks.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Table 4.8: Model selection criteria (DIC, DICc, WAIC and logarithmic score) and computational time (in minutes) from fitted models in the analysis of simulated data in the municipalities of Navarre and Basque Country. Simplified Laplace approximation.

	Spatially correlated one-dimensional temporal P-splines									
	$\mathbf{P}_k = \Delta_1'\Delta_1$					$\mathbf{P}_k = \Delta_2'\Delta_2$				
Interaction	DIC	DICc	WAIC	LS	time	DIC	DICc	WAIC	LS	time
Type I	50867	50928	50889	25453	7	50867	50928	50889	25453	8
Type II	50831	50881	50849	25431	11705	51170	51338	51199	25627	11355
Type III	50749	50785	50762	25385	20	50749	50790	50761	25386	19
Type IV	**50745**	**50770**	**50757**	**25381**	11475	51148	51329	51173	25616	8891
	Temporally correlated two-dimensional spatial P-splines									
Isotropic penalty	$\mathbf{P}_s = \lambda_s(\mathbf{I}_{k_2} \otimes \Delta_{d_1}'\Delta_{d_1} + \Delta_{d_2}'\Delta_{d_2} \otimes \mathbf{I}_{k_1})$									
	$\mathbf{R}_\gamma = \Delta_1'\Delta_1$					$\mathbf{R}_\gamma = \Delta_2'\Delta_2$				
Interaction	DIC	DICc	WAIC	LS	time	DIC	DICc	WAIC	LS	time
Type I	50980	51029	51013	25516	7	50979	51027	51012	25515	7
Type II	**50869**	**50888**	**50889**	**25447**	514	50922	50950	50944	25476	483
Type III	50949	50974	50979	25495	22	50947	50972	50977	25494	21
Type IV	50874	50888	50894	25449	770	50921	50949	50942	25475	742
Anisotropic penalty	$\mathbf{P}_s = \lambda_{s_1}(\mathbf{I}_{k_2} \otimes \Delta_{d_1}'\Delta_{d_1}) + \lambda_{s_2}(\Delta_{d_2}'\Delta_{d_2} \otimes \mathbf{I}_{k_1})$									
	$\mathbf{R}_\gamma = \Delta_1'\Delta_1$					$\mathbf{R}_\gamma = \Delta_2'\Delta_2$				
Interaction	DIC	DICc	WAIC	LS	time	DIC	DICc	WAIC	LS	time
Type I	50981	51030	51014	25517	9	50980	51028	51013	25516	9
Type II	**50870**	**50888**	**50890**	**25448**	804	50923	50951	50945	25476	818
Type III	50950	50971	50979	25494	53	50948	50968	50977	25493	35
Type IV	50874	50888	50894	25449	2193	50924	50952	50946	25477	2284
	Interaction P-spline model									
	$\Delta_{d_t} \equiv 1_{st}$ order penalty					$\Delta_{d_t} \equiv 2_{nd}$ order penalty				
	DIC	DICc	WAIC	LS	time	DIC	DICc	WAIC	LS	time
$\Delta_{d_1}, \Delta_{d_2} \equiv 1_{st}$ order penalty	50903	50914	50923	25463	9	**50901**	**50911**	**50921**	**25462**	9
$\Delta_{d_1}, \Delta_{d_2} \equiv 2_{nd}$ order penalty	50914	50925	50934	25468	13	50919	50928	50938	25470	14
	ANOVA-type P-spline model									
	$\Delta_{d_t} \equiv 1_{st}$ order penalty					$\Delta_{d_t} \equiv 2_{nd}$ order penalty				
	DIC	DICc	WAIC	LS	time	DIC	DICc	WAIC	LS	time
$\Delta_{d_1}, \Delta_{d_2} \equiv 1_{st}$ order penalty	50908	50914	50926	25464	68	**50890**	**50896**	**50906**	**25454**	69
$\Delta_{d_1}, \Delta_{d_2} \equiv 2_{nd}$ order penalty	50908	50914	50925	25463	83	50912	50917	50927	25464	67

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Depending on the prior distributions for the unknown regression coefficients, B-splines may become P-splines. The priors also account for temporal or spatial correlation. For the first class of models, random walk prior distributions can be placed over the temporal B-spline regression coefficients for each area giving rise to temporal P-splines, whereas a CAR prior can be also given to the coefficients of each time point according to the neighbourhood structure of the areas. If a normal prior with an identity precision matrix is used for the temporal coefficients, then B-splines models instead of P-splines are defined. For the second class of models, a normal prior with precision matrix based on first or second order differences in longitude and latitude leads to two-dimensional P-splines. This precision matrix may or may not account for anisotropy. If the identity is used as precision matrix, then two-dimensional B-splines appear. A random walk prior on the coefficients of the spatial surfaces in time would account for temporal correlation. Finally, for three-dimensional P-spline models, a normal prior with precision matrix based on first or second order difference matrices in longitude, latitude, and time is considered.

It is interesting to point out that the two-dimensional and three-dimensional splines deal with spatial dependence incorporating a smooth spatial surface in the mean, whereas the one-dimensional model includes the spatial effects through the covariance matrix. The main difference between these two approaches is that CAR spatial random effects produces local smoothing, while the two and three-dimensional P-splines lead, in principle, to large scale smoothing. This is so because splines take into account distance through the B-spline basis, which is constructed using some knots. On the other hand, the smoothing with CAR spatial random effects is local as they only consider adjacent areas (usually areas sharing a common border). However, the spatial smoothing with B-splines would be more similar to the CAR smoothing if the number of B-splines in the basis increases (increasing the number of knots). The reason is that if the number of B-splines is small, non adjacent regions will be covered with the same B-splines yielding to large scale smoothing. When analyzing real data we can proceed, in general, as follows: if we think there may be local factors as for example diet (which in Spain is rather local) affecting the disease, a CAR model could be the choice. If on the contrary there is a source of pollution vanishing smoothly with distance, a spline could be appropriate.

In this chapter, two and three-dimensional B-spline bases are based on tensor products of marginal bases. Tensor product smooths are very convenient for representing smooth interaction functions of several variables which are measured in different units. The key point is that if the scale is arbitrary, the smooth should be invariant to scaling and to arbitrary decisions about the relative smoothing with respect to the covariates. B-spline bases and tensor products of B-splines are invariant to linear transformation of covariates and smooths based on unpenalized tensor products of B-splines are scaling invariant (Wood et al., 2013). However, some cau-

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

tion is recommended with P-splines based on tensor product of B-splines as they may not be scaling invariant if penalties are not appropriately chosen. In particular, an isotropic penalty produces an arbitrary smoothing with respect to the covariates depending on the scale units. In our example this is not serious, as the scale range in longitude and latitude is similar. The three dimensional P-spline models considered in this paper (the pure interaction and the ANOVA-type) smooth covariates (space and time) measured in different units, but they are scaling invariant as different smoothing parameters are considered in each direction.

All the models are fitted from a fully Bayes approach using integrated nested Laplace approximations (INLA) to estimate the posterior marginal distributions. Anisotropic precision matrices can be fitted in `R-INLA` using the `generic3` model, allowing for different amount of smoothing in each direction. This is crucial as anisotropic models are scaling invariant whereas isotropic models, such as those fitted by Bauer et al. (2016) are not. As far as we know, this is the first work in which Bayesian three-dimensional P-splines models are fitted to smooth spatio-temporal count data. Additionally, special attention has been placed in the specification of the necessary identifiability constraints for each model. A correct specification of constrains is also relevant because placing more constraints than needed leads to restrictive functional shapes that may not recover the true risks. On the other hand, if less constraints than needed are considered, the model may crash or even confounding problems may arise inflating the posterior variance of some parameters. Details about which and how many constraints are required in the models considered in this chapter are provided.

The complete set of models has been used to analyze female breast cancer mortality data in Spanish provinces during the period 1990-2010 (Section 4.3.1). In this example, 47 small areas (Spanish provinces) are considered. To describe the performance of the models in a scenario with a higher number of small areas, an artificial data set based on the 501 municipalities of Navarre and Basque Country has been also analyzed (Section 4.3.2). In both scenarios, the model with spatially correlated one-dimensional temporal P-splines was selected as the best model in terms of all model selection criteria. This is an attractive model as it is easy to interpret: a temporal P-spline is fitted for each area and temporal P-splines from neighbouring regions tend to be similar. However, clear differences have been observed from a computational point of view. If a "small-to-moderate" number of small areas is analyzed (the number of areas is smaller than the number of columns in the two-dimensional B-spline basis, $n < k_1 k_2$), area specific temporal P-spline models will be the most appropriate computationally. Conversely, when the number of small areas is quite high, two-dimensional spatial P-spline models could be more suitable if the number of internal knots used for longitude and latitude is not so high (i.e., if the number of columns in the two-dimensional B-spline basis is smaller than the num-

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

ber of small areas, $k_1 k_2 < n$). In this case, temporal P-splines (Type II and Type IV interactions) are much slower than the two-dimensional B or P-splines. If Type I or Type III interactions are used, both modelling strategies are computationally similar and quick. In general, three-dimensional P-splines are always competitive models in terms of computing time, since the number of identifiability constraints is smaller than in the other models. Although they are not necessarily superior to the other model alternatives in terms of model selection criteria as we have seen in the two examples analyzed here, additional analyses with sparse data (not shown here) have shown the superiority of the ANOVA-type P-spline model with regard to model assessment criteria and speed of computations.

The contents of this chapter have been accepted for publication in Spatial Statistics.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Appendix 4A: Identifiability constraints for B-spline models

As in Appendix 3A, a spectral decomposition of the precision matrices of the B-splines coefficients is used to reparameterize the one, two and three-dimensional splines models described in Section 4.2, showing how the identifiability sum-to-zero constraints are derived. Full details for the models with completely structured (Type IV) interaction term are given below.

One-dimensional B-spline models

The one-dimensional B-spline model described in Equation (4.1) can be expressed in matrix form as

$$\log \mathbf{r} = (\mathbf{1}_n \otimes \mathbf{1}_T)\eta + (\mathbf{I}_n \otimes \mathbf{1}_T)\boldsymbol{\xi} + (\mathbf{1}_n \otimes \mathbf{B}_t)\boldsymbol{\theta}^{(t)} + (\mathbf{I}_n \otimes \mathbf{B}_t)\boldsymbol{\theta}^{(st)}$$

where $\mathbf{r} = (r_{11}, \ldots, r_{1T}, \ldots, r_{n1}, \ldots, r_{nT})^{'}$, $\mathbf{1}_n$ and $\mathbf{1}_T$ are vectors of ones of length n and T respectively, $\mathbf{I}_n$ is a $n \times n$ identity matrix and $\mathbf{B}_t$ is the temporal B-spline basis of dimension $T \times k_t$ obtained from the time covariate $\mathbf{x}_t = (x_1, \ldots, x_T)^{'}$. In Section 4.2.1, the following prior distributions are assumed for the random effects

$$\begin{aligned}
\boldsymbol{\xi} &\sim N(\mathbf{0}, [\tau_\xi \mathbf{D}_\xi]^{-1}); \quad \mathbf{D}_\xi = \lambda_\xi \mathbf{R}_\xi + (1-\lambda_\xi)\mathbf{I}_n, \\
\boldsymbol{\theta}^{(t)} &\sim N(\mathbf{0}, [\lambda_t \mathbf{P}_{k_t}]^{-}); \quad \mathbf{P}_{k_t} = \Delta_{d_t}^{'}\Delta_{d_t}, \quad \text{(RW1/RW2)} \\
\boldsymbol{\theta}^{(st)} &\sim N(\mathbf{0}, [\tau_{st}(\mathbf{R}_\xi \otimes \mathbf{P}_{k_t})]^{-}); \quad \text{(Type IV interaction)}.
\end{aligned}$$

where first or second order difference matrices can be considered for Δ_{d_t}.

If a RW1 prior distribution is assumed for the coefficients of the temporal B-splines (that is, temporal P-splines), let us consider the spectral decomposition of the structure matrices $\mathbf{D}_\xi$, $\mathbf{P}_{k_t}$ and $(\mathbf{R}_\xi \otimes \mathbf{P}_{k_t})$ as

$$\begin{aligned}
\mathbf{D}_\xi &= \mathbf{U}_\xi(\lambda_\xi \boldsymbol{\Sigma}_\xi + (1-\lambda_\xi)\mathbf{I}_n)\mathbf{U}_\xi^{'}, \\
\mathbf{P}_{k_t} &= \mathbf{U}_t \boldsymbol{\Sigma}_t \mathbf{U}_t^{'}, \\
\mathbf{R}_\xi \otimes \mathbf{P}_{k_t} &= \mathbf{U}_{st}(\boldsymbol{\Sigma}_\xi \otimes \boldsymbol{\Sigma}_t)\mathbf{U}_{st}^{'},
\end{aligned}$$

with

$$\mathbf{U}_\xi = [\mathbf{1}_n : \mathbf{U}_{\xi s}], \quad \mathbf{U}_t = [\mathbf{1}_{k_t} \mid \mathbf{U}_{\theta^{(t)}}],$$

$$\mathbf{U}_{st} = [\mathbf{1}_n \otimes \mathbf{1}_{k_t} : \mathbf{1}_n \otimes \mathbf{U}_{\theta^{(t)}} : \mathbf{U}_{\xi s} \otimes \mathbf{1}_{k_t} \mid \mathbf{U}_{\xi s} \otimes \mathbf{U}_{\theta^{(t)}}],$$

where $\mathbf{1}_n$ and $\mathbf{1}_{k_t}$ are vector of ones (up to a constant) of length n and k_t respectively,

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

$\mathbf{U}_{\xi s}$ is the $n \times (n-1)$ matrix whose columns are eigenvectors of $\mathbf{R}_\xi$ having non null-eigenvalues, $\mathbf{U}_{\theta^{(t)}}$ is the $k_t \times (k_t - 1)$ matrix whose columns are eigenvectors of $\mathbf{P}_{k_t}$ having non null-eigenvalues, $\mathbf{\Sigma}_\xi$ is a diagonal matrix with the eigenvalues of $\mathbf{R}_\xi$ in the main diagonal and $\mathbf{\Sigma}_t$ is a diagonal matrix with the eigenvalues of $\mathbf{P}_{k_t}$ in the main diagonal.

Since $\mathbf{U}_\xi$, $\mathbf{U}_t$ and $\mathbf{U}_{st}$ are orthogonal matrices, using the reparameterization described in Equation (1.5) we obtain that

$$\begin{aligned}
(\mathbf{I}_n \otimes \mathbf{1}_T)\boldsymbol{\xi} &= (\mathbf{I}_n \otimes \mathbf{1}_T)\mathbf{U}_\xi \mathbf{U}_\xi' \boldsymbol{\xi} = (\mathbf{I}_n \otimes \mathbf{1}_T)[\mathbf{1}_n : \mathbf{U}_{\xi s}] \begin{bmatrix} \mathbf{1}_n' \\ \mathbf{U}_{\xi s}' \end{bmatrix} \boldsymbol{\xi} \\
&= (\mathbf{1}_n \otimes \mathbf{1}_T)\beta_\xi + (\mathbf{U}_{\xi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_\xi, \\
(\mathbf{1}_n \otimes \mathbf{B}_t)\boldsymbol{\theta}^{(t)} &= (\mathbf{1}_n \otimes \mathbf{B}_t)\mathbf{U}_t \mathbf{U}_t' \boldsymbol{\theta}^{(t)} = (\mathbf{1}_n \otimes \mathbf{B}_t)[\mathbf{1}_{k_t} : \mathbf{U}_{\theta^{(t)}}] \begin{bmatrix} \mathbf{1}_{k_t}' \\ \mathbf{U}_{\theta^{(t)}}' \end{bmatrix} \boldsymbol{\theta}^{(t)} \\
&= (\mathbf{1}_n \otimes \mathbf{B}_t\mathbf{1}_{k_t})\beta_{\theta^{(t)}} + (\mathbf{1}_n \otimes \mathbf{B}_t\mathbf{U}_{\theta^{(t)}})\boldsymbol{\alpha}_{\theta^{(t)}}, \\
(\mathbf{I}_n \otimes \mathbf{B}_t)\boldsymbol{\theta}^{(st)} &= (\mathbf{I}_n \otimes \mathbf{B}_t)\mathbf{U}_{st}\mathbf{U}_{st}'\boldsymbol{\theta}^{(st)} \\
&= (\mathbf{I}_n \otimes \mathbf{B}_t)[\mathbf{1}_n \otimes \mathbf{1}_{k_t} : \mathbf{1}_n \otimes \mathbf{U}_{\theta^{(t)}} : \mathbf{U}_{\xi s} \otimes \mathbf{1}_{k_t} \mid \mathbf{U}_{\xi s} \otimes \mathbf{U}_{\theta^{(t)}}] \begin{bmatrix} \mathbf{1}_n' \otimes \mathbf{1}_{k_t}' \\ \mathbf{1}_n' \otimes \mathbf{U}_{\theta^{(t)}}' \\ \mathbf{U}_{\xi s}' \otimes \mathbf{1}_{k_t}' \\ \mathbf{U}_{\xi s}' \otimes \mathbf{U}_{\theta^{(t)}}' \end{bmatrix} \boldsymbol{\theta}^{(st)} \\
&= [\mathbf{1}_n \otimes \mathbf{B}_t\mathbf{1}_{k_t} : \mathbf{1}_n \otimes \mathbf{B}_t\mathbf{U}_{\theta^{(t)}} : \mathbf{U}_{\xi s} \otimes \mathbf{B}_t\mathbf{1}_{k_t}]\boldsymbol{\beta}_{\theta^{(st)}} + (\mathbf{U}_{\xi s} \otimes \mathbf{B}_t\mathbf{U}_{\theta^{(t)}})\boldsymbol{\alpha}_{\theta^{(st)}}.
\end{aligned}$$

Replacing these expressions into the model equation (note that $\mathbf{B}_t\mathbf{1}_{k_t} = \mathbf{1}_T$), the one-dimensional B-spline model can be reformulated as

$$\begin{aligned}
\log \mathbf{r} = {} & (\mathbf{1}_n \otimes \mathbf{1}_T)\eta + (\mathbf{1}_n \otimes \mathbf{1}_T)\boldsymbol{\beta}_\xi + (\mathbf{U}_{\xi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_\xi \\
& + (\mathbf{1}_n \otimes \mathbf{1}_T)\beta_{\theta^{(t)}} + (\mathbf{1}_n \otimes \mathbf{B}_t\mathbf{U}_{\theta^{(t)}})\boldsymbol{\alpha}_{\theta^{(t)}} \\
& + [\mathbf{1}_n \otimes \mathbf{1}_T : \mathbf{1}_n \otimes \mathbf{B}_t\mathbf{U}_{\theta^{(t)}} : \mathbf{U}_{\xi s} \otimes \mathbf{1}_T]\boldsymbol{\beta}_{\theta^{(st)}} + (\mathbf{U}_{\xi s} \otimes \mathbf{B}_t\mathbf{U}_{\theta^{(t)}})\boldsymbol{\alpha}_{\theta^{(st)}},
\end{aligned}$$

with

$$\boldsymbol{\alpha}_\xi \sim N(\mathbf{0}, [\tau_\xi(\lambda_\xi \widetilde{\mathbf{\Sigma}}_\xi + (1-\lambda_\xi)\mathbf{I}_{n-1})]^{-1}),$$

$$\boldsymbol{\alpha}_{\theta^{(t)}} \sim N(\mathbf{0}, [\lambda_t \widetilde{\mathbf{\Sigma}}_t]^{-1}) \quad \text{and} \quad \boldsymbol{\alpha}_{\theta^{(st)}} \sim N(\mathbf{0}, [\tau_{st}(\widetilde{\mathbf{\Sigma}}_\xi \otimes \widetilde{\mathbf{\Sigma}}_t)]^{-1}),$$

where $\widetilde{\mathbf{\Sigma}}_\xi$ and $\widetilde{\mathbf{\Sigma}}_t$ are diagonal matrices with the non-null eigenvalues of $\mathbf{R}_\xi$, and $\mathbf{P}_{k_t}$ respectively. If we remove the repeated columns $\mathbf{1}_n \otimes \mathbf{1}_T$ (corresponding to β_ξ, $\beta_{\theta^{(t)}}$ and $\boldsymbol{\beta}_{\theta^{(st)}}$), $\mathbf{1}_n \otimes \mathbf{B}_t\mathbf{U}_{\theta^{(t)}}$ and $\mathbf{U}_{\xi s} \otimes \mathbf{1}_T$ (corresponding to $\boldsymbol{\beta}_{\theta^{(st)}}$), this leaves the following model

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

$$\log \mathbf{r} = (\mathbf{1}_n \otimes \mathbf{1}_T)\eta + (\mathbf{U}_{\xi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_\xi + (\mathbf{1}_n \otimes \mathbf{B}_t\mathbf{U}_{\theta^{(t)}})\boldsymbol{\alpha}_{\theta^{(t)}} + (\mathbf{U}_{\xi s} \otimes \mathbf{B}_t\mathbf{U}_{\theta^{(t)}})\boldsymbol{\alpha}_{\theta^{(st)}}.$$

Removing (or setting to zero) the repeated columns makes the model identifiable, and now the precision matrices of the reparameterized random effects have full rank. Note that removing the repeated columns leads to the linear constraints (see Goicoa et al., 2017 for details)

$$\sum_{i=1}^{n} \xi_i = 0; \quad \sum_{k=1}^{k_t} \theta_k = 0; \quad \begin{array}{l} \sum_{k=1}^{k_t} \theta_{ik}^{(st)} = 0, \quad \text{for} \quad i = 1, \ldots, n \\ \sum_{i=1}^{n} \theta_{ik}^{(st)} = 0, \quad \text{for} \quad k = 1, \ldots, k_t \end{array}$$

On the other hand, if a RW2 prior distribution is assumed for the coefficients of the temporal P-splines (that is, $\Delta_{d_t} \equiv 2_{nd}$ order difference matrix), the orthogonal matrices $\mathbf{U}_\xi$, $\mathbf{U}_t$ and $\mathbf{U}_{st}$ are now expressed as

$$\mathbf{U}_\xi = [\mathbf{1}_n : \mathbf{U}_{\xi s}], \quad \mathbf{U}_t = [\mathbf{1}_{k_t} : \mathbf{k}_t^* \mid \mathbf{U}_{\theta^{(t)}}],$$

$$\mathbf{U}_{st} = [\mathbf{1}_n \otimes \mathbf{1}_{k_t} : \mathbf{1}_n \otimes \mathbf{k}_t^* : \mathbf{1}_n \otimes \mathbf{U}_{\theta^{(t)}} : \mathbf{U}_{\xi s} \otimes \mathbf{1}_{k_t} : \mathbf{U}_{\xi s} \otimes \mathbf{k}_t^* \mid \mathbf{U}_{\xi s} \otimes \mathbf{U}_{\theta^{(t)}}],$$

where $\mathbf{k}_t^* = (1, 2, \ldots, k_t)'$ (up to a normalizing constant) and $\mathbf{U}_{\theta^{(t)}}$ is the $k_t \times (k_t - 2)$ matrix whose columns are eigenvectors of $\mathbf{P}_{k_t}$ having non null-eigenvalues. In this case, reparameterizing the random effects as described above and removing the repeated columns of the design matrices, leads to the model

$$\begin{aligned} \log \mathbf{r} = {} & (\mathbf{1}_n \otimes \mathbf{1}_T)\eta + (\mathbf{1}_n \otimes \mathbf{B}_t\mathbf{k}_t^*)\beta_{\theta^{(t)}} + (\mathbf{U}_{\xi s} \otimes \mathbf{B}_t\mathbf{k}_t^*)\beta_{\theta^{(st)}} \\ & + (\mathbf{U}_{\xi s} \otimes \mathbf{1}_T)\boldsymbol{\alpha}_\xi + (\mathbf{1}_n \otimes \mathbf{B}_t\mathbf{U}_{\theta^{(t)}})\boldsymbol{\alpha}_{\theta^{(t)}} + (\mathbf{U}_{\xi s} \otimes \mathbf{B}_t\mathbf{U}_{\theta^{(t)}})\boldsymbol{\alpha}_{\theta^{(st)}}. \end{aligned}$$

It can be shown that removing the repeated terms lead to the same sum-to-zero constraints as if a RW1 prior distribution is considered for the temporal P-spline coefficients.

Two-dimensional B-spline models

The two-dimensional B-spline model described in Equation (4.2) can be expressed in matrix form as

$$\log \mathbf{r} = (\mathbf{1}_T \otimes \mathbf{1}_n)\eta + (\mathbf{1}_T \otimes \mathbf{B}_s)\boldsymbol{\theta}^{(s)} + (\mathbf{I}_T \otimes \mathbf{1}_n)\gamma + (\mathbf{I}_T \otimes \mathbf{B}_s)\boldsymbol{\theta}^{(st)}$$

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

where $\mathbf{r} = (r_{11}, \ldots, r_{n1}, \ldots, r_{1T}, \ldots, r_{nT})'$, $\mathbf{1}_n$ and $\mathbf{1}_T$ are vectors of ones of length n and T respectively, $\mathbf{I}_T$ is a $T \times T$ identity matrix and $\mathbf{B}_s$ is the spatial two-dimensional B-spline basis of dimension $n \times k_1 k_2$ obtained from the row-wise Kronecker product of the marginal bases for longitude $\mathbf{x}_1 = (x_{11}, \ldots, x_{1n})'$ and latitude $\mathbf{x}_2 = (x_{21}, \ldots, x_{2n})'$. In Section 4.2.2, the following prior distributions are assumed for the random effects

$$\begin{aligned} \boldsymbol{\theta}^{(s)} &\sim N(\mathbf{0}, \mathbf{P}_s^-); \quad \mathbf{P}_s = \lambda_s(\mathbf{I}_{k_2} \otimes \Delta_{d_1}'\Delta_{d_1} + \Delta_{d_2}'\Delta_{d_2}\mathbf{I}_{k_1}), \\ \boldsymbol{\gamma} &\sim N(\mathbf{0}, [\tau_\gamma \mathbf{R}_\gamma]^-); \quad \mathbf{R}_\gamma = \Delta_{d_t}'\Delta_{d_t}, \quad \text{(RW1/RW2)} \\ \boldsymbol{\theta}^{(st)} &\sim N(\mathbf{0}, [\tau_{st}(\mathbf{R}_\gamma \otimes \mathbf{P}_s)]^-); \quad \text{(Type IV interaction)} \end{aligned}$$

where first or second order difference matrices can be considered for Δ_{d_1}, Δ_{d_2} and Δ_{d_t}. Note that both the two-dimensional anisotropic and isotropic penalty matrices $\mathbf{P}_s$ (defined in Equation (4.3) and Equation (4.4) respectively) have the same null and non-null eigenvectors but corresponding to different eigenvalues. For notation simplicity, the isotropic penalty has been considered here.

If first order difference matrices are considered for the precision matrix of the spatial two-dimensional B-spline (that is, spatial P-splines), let us consider the spectral decomposition of the structure matrices $\mathbf{P}_s$, $\mathbf{R}_\gamma$ and $(\mathbf{R}_\gamma \otimes \mathbf{P}_s)$ as

$$\begin{aligned} \mathbf{P}_s &= \mathbf{U}_s \boldsymbol{\Sigma}_s \mathbf{U}_s', \\ \mathbf{R}_\gamma &= \mathbf{U}_\gamma \boldsymbol{\Sigma}_\gamma \mathbf{U}_\gamma', \\ \mathbf{R}_\gamma \otimes \mathbf{P}_s &= \mathbf{U}_{st}(\boldsymbol{\Sigma}_\gamma \otimes \boldsymbol{\Sigma}_s)\mathbf{U}_{st}', \end{aligned}$$

with

$$\mathbf{U}_s = [\mathbf{1}_{k_1 k_2} : \mathbf{U}_{\theta^{(s)}}], \quad \mathbf{U}_\gamma = [\mathbf{1}_T \mid \mathbf{U}_{\gamma s}],$$

$$\mathbf{U}_{st} = [\mathbf{1}_T \otimes \mathbf{1}_{k_1 k_2} : \mathbf{1}_T \otimes \mathbf{U}_{\theta^{(s)}} : \mathbf{U}_{\gamma s} \otimes \mathbf{1}_{k_1 k_2} \mid \mathbf{U}_{\gamma s} \otimes \mathbf{U}_{\theta^{(s)}}],$$

where $\mathbf{1}_{k_1 k_2}$ and $\mathbf{1}_T$ are vector of ones (up to a constant) of length $k_1 k_2$ and T respectively, $\mathbf{U}_{\theta^{(s)}}$ is the $k_1 k_2 \times (k_1 k_2 - 1)$ matrix whose columns are eigenvectors of $\mathbf{P}_s$ having non null-eigenvalues, $\mathbf{U}_{\gamma s}$ is the $T \times (T-1)$ matrix whose columns are eigenvectors of $\mathbf{R}_\gamma$ having non null-eigenvalues, $\boldsymbol{\Sigma}_s$ is a diagonal matrix with the eigenvalues of $\mathbf{P}_s$ in the main diagonal and $\boldsymbol{\Sigma}_\gamma$ is a diagonal matrix with the eigenvalues of $\mathbf{R}_\gamma$ in the main diagonal.

Since $\mathbf{U}_s$, $\mathbf{U}_\gamma$ and $\mathbf{U}_{st}$ are orthogonal matrices, using the reparameterization described in Equation (1.5) we obtain that

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

$$
\begin{aligned}
(\mathbf{1}_T \otimes \mathbf{B}_s)\boldsymbol{\theta}^{(s)} &= (\mathbf{1}_T \otimes \mathbf{B}_s)\mathbf{U}_s\mathbf{U}_s'\boldsymbol{\theta}^{(s)} = (\mathbf{1}_T \otimes \mathbf{B}_s)[\mathbf{1}_{k_1k_2} : \mathbf{U}_{\theta^{(s)}}] \begin{bmatrix} \mathbf{1}_{k_1k_2}' \\ \mathbf{U}_{\theta^{(s)}}' \end{bmatrix} \boldsymbol{\theta}^{(s)} \\
&= (\mathbf{1}_T \otimes \mathbf{B}_s\mathbf{1}_{k_1k_2})\beta_{\theta^{(s)}} + (\mathbf{1}_T \otimes \mathbf{B}_s\mathbf{U}_{\theta^{(s)}})\boldsymbol{\alpha}_{\theta^{(s)}},
\end{aligned}
$$

$$
\begin{aligned}
(\mathbf{I}_T \otimes \mathbf{1}_n)\boldsymbol{\gamma} &= (\mathbf{I}_T \otimes \mathbf{1}_n)\mathbf{U}_\gamma\mathbf{U}_\gamma'\boldsymbol{\gamma} = (\mathbf{I}_T \otimes \mathbf{1}_n)[\mathbf{1}_T : \mathbf{U}_{\gamma s}] \begin{bmatrix} \mathbf{1}_T' \\ \mathbf{U}_{\gamma s}' \end{bmatrix} \boldsymbol{\gamma} \\
&= (\mathbf{1}_T \otimes \mathbf{1}_n)\beta_\gamma + (\mathbf{U}_{\gamma s} \otimes \mathbf{1}_n)\boldsymbol{\alpha}_\gamma,
\end{aligned}
$$

$$
\begin{aligned}
(\mathbf{I}_T \otimes \mathbf{B}_s)\boldsymbol{\theta}^{(st)} &= (\mathbf{I}_T \otimes \mathbf{B}_s)\mathbf{U}_{st}\mathbf{U}_{st}'\boldsymbol{\theta}^{(st)} \\
&= (\mathbf{I}_T \otimes \mathbf{B}_s)[\mathbf{1}_T \otimes \mathbf{1}_{k_1k_2} : \mathbf{1}_T \otimes \mathbf{U}_{\theta^{(s)}} : \mathbf{U}_{\gamma s} \otimes \mathbf{1}_{k_1k_2} \mid \mathbf{U}_{\gamma s} \otimes \mathbf{U}_{\theta^{(s)}}] \begin{bmatrix} \mathbf{1}_T' \otimes \mathbf{1}_{k_1k_2}' \\ \mathbf{1}_T' \otimes \mathbf{U}_{\theta^{(s)}}' \\ \mathbf{U}_{\gamma s}' \otimes \mathbf{1}_{k_1k_2}' \\ \mathbf{U}_{\gamma s}' \otimes \mathbf{U}_{\theta^{(s)}}' \end{bmatrix} \boldsymbol{\theta}^{(st)} \\
&= [\mathbf{1}_T \otimes \mathbf{B}_s\mathbf{1}_{k_1k_2} : \mathbf{1}_T \otimes \mathbf{B}_s\mathbf{U}_{\theta^{(s)}} : \mathbf{U}_{\gamma s} \otimes \mathbf{B}_s\mathbf{1}_{k_1k_2}]\boldsymbol{\beta}_{\theta^{(st)}} + (\mathbf{U}_{\gamma s} \otimes \mathbf{B}_s\mathbf{U}_{\theta^{(s)}})\boldsymbol{\alpha}_{\theta^{(st)}}.
\end{aligned}
$$

Replacing these expressions into the model equation (note that $\mathbf{B}_s\mathbf{1}_{k_1k_2} = \mathbf{1}_n$), the two-dimensional B-spline model can be reformulated as

$$
\begin{aligned}
\log \mathbf{r} = {} & (\mathbf{1}_T \otimes \mathbf{1}_n)\eta + (\mathbf{1}_T \otimes \mathbf{1}_n)\boldsymbol{\beta}_{\theta^{(s)}} + (\mathbf{1}_T \otimes \mathbf{B}_s\mathbf{U}_{\theta^{(s)}})\boldsymbol{\alpha}_{\theta^{(s)}} \\
& + (\mathbf{1}_T \otimes \mathbf{1}_n)\beta_\gamma + (\mathbf{U}_{\gamma s} \otimes \mathbf{1}_n)\boldsymbol{\alpha}_\gamma \\
& + [\mathbf{1}_T \otimes \mathbf{B}_s\mathbf{1}_{k_1k_2} : \mathbf{1}_T \otimes \mathbf{B}_s\mathbf{U}_{\theta^{(s)}} : \mathbf{U}_{\gamma s} \otimes \mathbf{B}_s\mathbf{1}_{k_1k_2}]\boldsymbol{\beta}_{\theta^{(st)}} + (\mathbf{U}_{\gamma s} \otimes \mathbf{B}_s\mathbf{U}_{\theta^{(s)}})\boldsymbol{\alpha}_{\theta^{(st)}},
\end{aligned}
$$

with

$$
\boldsymbol{\alpha}_{\theta^{(s)}} \sim N(\mathbf{0}, [\lambda_s\widetilde{\boldsymbol{\Sigma}}_s]^{-1}), \quad \boldsymbol{\alpha}_\gamma \sim N(\mathbf{0}, [\tau_\gamma\widetilde{\boldsymbol{\Sigma}}_\gamma]^{-1}),
$$

$$
\boldsymbol{\alpha}_{\theta^{(st)}} \sim N(\mathbf{0}, [\tau_{st}(\widetilde{\boldsymbol{\Sigma}}_\gamma \otimes \widetilde{\boldsymbol{\Sigma}}_s)]^{-1}),
$$

where $\widetilde{\boldsymbol{\Sigma}}_s$ and $\widetilde{\boldsymbol{\Sigma}}_\gamma$ are diagonal matrices with the non-null eigenvalues of $\mathbf{P}_s$, and $\mathbf{R}_\gamma$ respectively. If we remove the repeated columns $\mathbf{1}_T \otimes \mathbf{1}_n$ (corresponding to $\beta_{\theta^{(s)}}$, β_γ and $\boldsymbol{\beta}_{\theta^{(st)}}$), $\mathbf{1}_T \otimes \mathbf{B}_s\mathbf{U}_{\theta^{(s)}}$ and $\mathbf{U}_{\gamma s} \otimes \mathbf{1}_n$ (corresponding to $\boldsymbol{\beta}_{\theta^{(st)}}$), this leaves the following model

$$
\log \mathbf{r} = (\mathbf{1}_T \otimes \mathbf{1}_n)\eta + (\mathbf{1}_T \otimes \mathbf{B}_s\mathbf{U}_{\theta^{(s)}})\boldsymbol{\alpha}_{\theta^{(s)}} + (\mathbf{U}_{\gamma s} \otimes \mathbf{1}_n)\boldsymbol{\alpha}_\gamma + (\mathbf{U}_{\gamma s} \otimes \mathbf{B}_s\mathbf{U}_{\theta^{(s)}})\boldsymbol{\alpha}_{\theta^{(st)}}.
$$

Removing (or setting to zero) the repeated columns makes the model identifiable, and now the precision matrices of the reparameterized random effects have full rank. Note that removing the repeated columns leads to the linear constraints (see Goicoa et al., 2017 for details)

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

$$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)}=0; \quad \sum_{t=1}^{T}\gamma_t=0; \quad \begin{array}{l}\sum_{t=1}^{T}\theta_{ij,t}^{(st)}=0, \quad \text{for} \quad \begin{array}{l} i=1,\ldots,k_1 \\ j=1,\ldots,k_2\end{array} \\ \sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij,t}^{(st)}=0, \quad \text{for} \quad t=1,\ldots,T\end{array}$$

On the other hand, if second order difference matrices are considered for the coefficients of the two-dimensional spatial P-spline (that is, $\Delta_{d_1},\Delta_{d_2}\equiv 2_{nd}$ order difference matrices), the orthogonal matrices $\mathbf{U}_s$, $\mathbf{U}_\gamma$ and $\mathbf{U}_{st}$ are now expressed as

$$\begin{aligned}\mathbf{U}_s = \;& [\mathbf{1}_{k_1k_2} : \mathbf{1}_{k_2}\otimes\mathbf{k}_1^* : \mathbf{k}_2^*\otimes\mathbf{1}_{k_1} : \mathbf{k}_2^*\otimes\mathbf{k}_1^* \mid \mathbf{U}_{\theta^{(s)}}], \quad \mathbf{U}_\gamma=[\mathbf{1}_T \mid \mathbf{U}_{\gamma s}],\\ \mathbf{U}_{st} = \;& [\mathbf{1}_T\otimes\mathbf{1}_{k_1k_2} : \mathbf{1}_T\otimes\mathbf{1}_{k_2}\otimes\mathbf{k}_1^* : \mathbf{1}_T\otimes\mathbf{k}_2^*\otimes\mathbf{1}_{k_1} : \mathbf{1}_T\otimes\mathbf{k}_2^*\otimes\mathbf{k}_1^* :\\ & \mathbf{U}_{\gamma s}\otimes\mathbf{1}_{k_1k_2} : \mathbf{U}_{\gamma s}\otimes\mathbf{1}_{k_2}\otimes\mathbf{k}_1^* : \mathbf{U}_{\gamma s}\otimes\mathbf{k}_2^*\otimes\mathbf{1}_{k_1} : \mathbf{U}_{\gamma s}\otimes\mathbf{k}_2^*\otimes\mathbf{k}_1^* :\\ & \mathbf{1}_T\otimes\mathbf{U}_{\theta^{(s)}} \mid \mathbf{U}_{\gamma s}\otimes\mathbf{U}_{\theta^{(s)}}],\end{aligned}$$

where $\mathbf{k}_1^*=(1,2,\ldots,k_1)'$ and $\mathbf{k}_2^*=(1,2,\ldots,k_2)'$ (up to a normalizing constant), and $\mathbf{U}_{\theta^{(s)}}$ is the $k_1k_2\times(k_1k_2-4)$ matrix whose columns are eigenvectors of $\mathbf{P}_s$ having non null-eigenvalues. In this case, reparameterizing the random effects as described above and removing the repeated columns of the design matrices, leads to the model

$$\begin{aligned}\log\mathbf{r} = & (\mathbf{1}_T\otimes\mathbf{1}_n)\eta + [\mathbf{1}_T\otimes\mathbf{B}_s(\mathbf{1}_{k_2}\otimes\mathbf{k}_1^*) : \mathbf{1}_T\otimes\mathbf{B}_s(\mathbf{k}_2^*\otimes\mathbf{1}_{k_1}) : \mathbf{1}_T\otimes\mathbf{B}_s(\mathbf{k}_2^*\otimes\mathbf{k}_1^*)]\boldsymbol{\beta}_{\theta^{(s)}}\\ & + [\mathbf{U}_{\gamma s}\otimes\mathbf{B}_s(\mathbf{1}_{k_2}\otimes\mathbf{k}_1^*) : \mathbf{U}_{\gamma s}\otimes\mathbf{B}_s(\mathbf{k}_2^*\otimes\mathbf{1}_{k_1}) : \mathbf{U}_{\gamma s}\otimes\mathbf{B}_s(\mathbf{k}_2^*\otimes\mathbf{k}_1^*)]\boldsymbol{\beta}_{\theta^{(st)}}\\ & + (\mathbf{1}_T\otimes\mathbf{B}_s\mathbf{U}_{\theta^{(s)}})\boldsymbol{\alpha}_{\theta^{(s)}} + (\mathbf{U}_{\gamma s}\otimes\mathbf{1}_n)\boldsymbol{\alpha}_\gamma + (\mathbf{U}_{\gamma s}\otimes\mathbf{B}_s\mathbf{U}_{\theta^{(s)}})\boldsymbol{\alpha}_{\theta^{(st)}}.\end{aligned}$$

Again, it can be shown that removing the repeated terms lead to the same sum-to-zero constraints as if first order difference matrices are considered for the spatial P-spline coefficients.

Three-dimensional B-spline models

The ANOVA-type P-spline model described in Equation (2.3) can be expressed in matrix form as

$$\log\mathbf{r}=(\mathbf{1}_T\otimes\mathbf{1}_n)\eta+(\mathbf{1}_T\otimes\mathbf{B}_s)\boldsymbol{\theta}^{(s)}+(\mathbf{B}_t\otimes\mathbf{1}_n)\boldsymbol{\theta}^{(t)}+(\mathbf{B}_t\otimes\mathbf{B}_s)\boldsymbol{\theta}^{(st)}$$

where $\mathbf{r}=(r_{11},\ldots,r_{n1},\ldots,r_{1T},\ldots,r_{nT})'$, $\mathbf{1}_n$ and $\mathbf{1}_T$ are vectors of ones of length n and T respectively, $\mathbf{B}_t$ is the temporal B-spline basis of dimension $T\times k_t$ ob-

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

tained from the time covariate $\mathbf{x}_t = (x_1, \ldots, x_T)'$ and $\mathbf{B}_s$ is the spatial two-dimensional B-spline basis of dimension $n \times k_1 k_2$ obtained from the row-wise Kronecker product of the marginal bases for longitude $\mathbf{x}_1 = (x_{11}, \ldots, x_{1n})'$ and latitude $\mathbf{x}_2 = (x_{21}, \ldots, x_{2n})'$. In Section 4.2.3, the following prior distributions are assumed for the random effects

$$\boldsymbol{\theta}^{(s)} \sim N(\mathbf{0}, \mathbf{P}_s^-), \quad \boldsymbol{\theta}^{(t)} \sim N(\mathbf{0}, \mathbf{P}_{k_t}^-) \quad \text{and} \quad \boldsymbol{\theta}^{(st)} \sim N(\mathbf{0}, \mathbf{P}_{st}^-),$$

where

$$\mathbf{P}_s = \lambda_{s_1}(\mathbf{I}_{k_2} \otimes \Delta_{d_1}'\Delta_{d_1}) + \lambda_{s_2}(\Delta_{d_2}'\Delta_{d_2} \otimes \mathbf{I}_{k_1}), \quad \mathbf{P}_{k_t} = \lambda_t \Delta_{d_t}'\Delta_{d_t},$$

$$\mathbf{P}_{st} = \tau_1(\mathbf{I}_{k_t} \otimes \mathbf{I}_{k_2} \otimes \Delta_{d_1}'\Delta_{d_1}) + \tau_2(\mathbf{I}_{k_t} \otimes \Delta_{d_2}'\Delta_{d_2} \otimes \mathbf{I}_{k_1}) + \tau_t(\Delta_{d_t}'\Delta_{d_t} \otimes \mathbf{I}_{k_2} \otimes \mathbf{I}_{k_1}).$$

As for the previous models, first or second order difference matrices can be considered for Δ_{d_1}, Δ_{d_2} and Δ_{d_t}. If first order difference matrices are considered, the spectral decomposition of the precision matrices $\mathbf{P}_s$, $\mathbf{P}_{k_t}$ and $\mathbf{P}_{st}$ are expressed as

$$\mathbf{P}_s = \mathbf{U}_s \boldsymbol{\Sigma}_s \mathbf{U}_s', \quad \mathbf{P}_{k_t} = \mathbf{U}_t \boldsymbol{\Sigma}_t \mathbf{U}_t' \quad \text{and} \quad \mathbf{P}_{st} = \mathbf{U}_{st} \boldsymbol{\Sigma}_{st} \mathbf{U}_{st}',$$

with

$$\mathbf{U}_s = [\mathbf{1}_{k_1 k_2} \mid \mathbf{U}_{\theta^{(s)}}], \quad \mathbf{U}_t = [\mathbf{1}_{k_t} \mid \mathbf{U}_{\theta^{(t)}}], \quad \text{and} \quad \mathbf{U}_{st} = [\mathbf{1}_{k_t} \otimes \mathbf{1}_{k_1 k_2} \mid \mathbf{U}_{\theta^{(st)}}],$$

where $\mathbf{1}_{k_1 k_2}$ and $\mathbf{1}_{k_t}$ are vector of ones (up to a constant) of length $k_1 k_2$ and k_t respectively, $\mathbf{U}_{\theta^{(s)}}$ is the $k_1 k_2 \times (k_1 k_2 - 1)$ matrix whose columns are eigenvectors of $\mathbf{P}_s$ having non null-eigenvalues, $\mathbf{U}_{\theta^{(t)}}$ is the $k_t \times (k_t - 1)$ matrix whose columns are eigenvectors of $\mathbf{P}_{k_t}$ having non null-eigenvalues, $\mathbf{U}_{\theta^{(st)}}$ is the $k_1 k_2 k_t \times (k_1 k_2 k_t - 1)$ matrix whose columns are eigenvectors of $\mathbf{P}_{st}$ having non null-eigenvalues, while $\boldsymbol{\Sigma}_s$, $\boldsymbol{\Sigma}_t$ and $\boldsymbol{\Sigma}_{st}$ are diagonal matrices with the eigenvalues of $\mathbf{P}_s$, $\mathbf{P}_{k_t}$ and $\mathbf{P}_{st}$ in the main diagonal respectively.

Since $\mathbf{U}_s$, $\mathbf{U}_t$ and $\mathbf{U}_{st}$ are orthogonal matrices, using the reparameterization described in Equation (1.5) we obtain that

$$\begin{aligned}
(\mathbf{1}_T \otimes \mathbf{B}_s)\boldsymbol{\theta}^{(s)} &= (\mathbf{1}_T \otimes \mathbf{B}_s)\mathbf{U}_s\mathbf{U}_s'\boldsymbol{\theta}^{(s)} = (\mathbf{1}_T \otimes \mathbf{B}_s)[\mathbf{1}_{k_1 k_2} : \mathbf{U}_{\theta^{(s)}}] \begin{bmatrix} \mathbf{1}_{k_1 k_2}' \\ \mathbf{U}_{\theta^{(s)}}' \end{bmatrix} \boldsymbol{\theta}^{(s)} \\
&= (\mathbf{1}_T \otimes \mathbf{B}_s\mathbf{1}_{k_1 k_2})\beta_{\theta^{(s)}} + (\mathbf{1}_T \otimes \mathbf{B}_s\mathbf{U}_{\theta^{(s)}})\boldsymbol{\alpha}_{\theta^{(s)}},
\end{aligned}$$

$$\begin{aligned}
(\mathbf{B}_t \otimes \mathbf{1}_n)\boldsymbol{\theta}^{(t)} &= (\mathbf{B}_t \otimes \mathbf{1}_n)\mathbf{U}_t\mathbf{U}_t'\boldsymbol{\theta}^{(t)} = (\mathbf{B}_t \otimes \mathbf{1}_n)[\mathbf{1}_{k_t} : \mathbf{U}_{\theta^{(t)}}] \begin{bmatrix} \mathbf{1}_{k_t}' \\ \mathbf{U}_{\theta^{(t)}}' \end{bmatrix} \boldsymbol{\theta}^{(t)} \\
&= (\mathbf{B}_t\mathbf{1}_{k_t} \otimes \mathbf{1}_n)\beta_{\theta^{(t)}} + (\mathbf{B}_t\mathbf{U}_{\theta^{(t)}} \otimes \mathbf{1}_n)\boldsymbol{\alpha}_{\theta^{(t)}},
\end{aligned}$$

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

$$
\begin{aligned}
(\mathbf{B}_t \otimes \mathbf{B}_s)\boldsymbol{\theta}^{(st)} &= (\mathbf{B}_t \otimes \mathbf{B}_s)\mathbf{U}_{st}\mathbf{U}_{st}^{'}\boldsymbol{\theta}^{(st)} \\
&= (\mathbf{B}_t \otimes \mathbf{B}_s)[\mathbf{1}_{k_t} \otimes \mathbf{1}_{k_1k_2} : \mathbf{U}_{\theta^{(st)}}] \begin{bmatrix} \mathbf{1}_{k_t}^{'} \otimes \mathbf{1}_{k_1k_2}^{'} \\ \mathbf{U}_{\theta^{(st)}}^{'} \end{bmatrix} \boldsymbol{\theta}^{(st)} \\
&= (\mathbf{B}_t\mathbf{1}_{k_t} \otimes \mathbf{B}_s\mathbf{1}_{k_1k_2})\beta_{\theta^{(st)}} + [(\mathbf{B}_t \otimes \mathbf{B}_s)\mathbf{U}_{\theta^{(st)}}]\boldsymbol{\alpha}_{\theta^{(st)}}.
\end{aligned}
$$

Replacing these expressions into the model equation (note that $\mathbf{B}_t\mathbf{1}_{k_t} = \mathbf{1}_T$ and $\mathbf{B}_s\mathbf{1}_{k_1k_2} = \mathbf{1}_n$) and removing the repeated columns corresponding to $\beta_{\theta^{(s)}}$, $\beta_{\theta^{(t)}}$ and $\beta_{\theta^{(st)}}$, the ANOVA-type P-spline model is reformulated as

$$
\log \mathbf{r} = (\mathbf{1}_T \otimes \mathbf{1}_n)\eta + (\mathbf{1}_T \otimes \mathbf{B}_s\mathbf{U}_{\theta^{(s)}})\boldsymbol{\alpha}_{\theta^{(s)}} + (\mathbf{B}_t\mathbf{U}_{\theta^{(t)}} \otimes \mathbf{1}_n)\boldsymbol{\alpha}_{\theta^{(t)}} + [(\mathbf{B}_t \otimes \mathbf{B}_s)\mathbf{U}_{\theta^{(st)}}]\boldsymbol{\alpha}_{\theta^{(st)}},
$$

with

$$
\boldsymbol{\alpha}_{\theta^{(s)}} \sim N(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}}_s^{-1}), \quad \boldsymbol{\alpha}_{\theta^{(t)}} \sim N(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}}_t^{-1}) \quad \text{and} \quad \boldsymbol{\alpha}_{\theta^{(st)}} \sim N(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}}_{st}^{-1}),
$$

where $\widetilde{\boldsymbol{\Sigma}}_s$, $\widetilde{\boldsymbol{\Sigma}}_t$ and $\widetilde{\boldsymbol{\Sigma}}_{st}$ are diagonal matrices with the non-null eigenvalues of $\mathbf{P}_s$, $\mathbf{P}_{k_t}$ and $\mathbf{P}_{st}$ respectively. Note that removing the repeated columns leads to the linear constraints

$$
\boxed{\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)} = 0; \quad \sum_{k=1}^{k_t}\theta_{k}^{(t)} = 0; \quad \sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t}\theta_{ijk}^{(st)} = 0.}
$$

On the other hand, if second order difference matrices are considered for Δ_{d_1}, Δ_{d_2} and Δ_{d_t} the orthogonal matrices $\mathbf{U}_s$, $\mathbf{U}_t$ and $\mathbf{U}_{st}$ are now expressed as

$$
\begin{aligned}
\mathbf{U}_s = \; & [\mathbf{1}_{k_1k_2} : \mathbf{1}_{k_2} \otimes \mathbf{k}_1^* : \mathbf{k}_2^* \otimes \mathbf{1}_{k_1} : \mathbf{k}_2^* \otimes \mathbf{k}_1^* \mid \mathbf{U}_{\theta^{(s)}}], \quad \mathbf{U}_t = [\mathbf{1}_{k_t} : \mathbf{k}_t^* \mid \mathbf{U}_{\theta^{(s)}}], \\
\mathbf{U}_{st} = \; & [\mathbf{1}_{k_t} \otimes \mathbf{1}_{k_2} \otimes \mathbf{1}_{k_1} : \mathbf{1}_{k_t} \otimes \mathbf{1}_{k_2} \otimes \mathbf{k}_1^* : \mathbf{1}_{k_t} \otimes \mathbf{k}_2^* \otimes \mathbf{1}_{k_1} : \mathbf{1}_{k_t} \otimes \mathbf{k}_2^* \otimes \mathbf{k}_1^* : \\
& \mathbf{k}_t^* \otimes \mathbf{1}_{k_2} \otimes \mathbf{1}_{k_1} : \mathbf{k}_t^* \otimes \mathbf{1}_{k_2} \otimes \mathbf{k}_1^* : \mathbf{k}_t^* \otimes \mathbf{k}_2^* \otimes \mathbf{1}_{k_1} : \mathbf{k}_t^* \otimes \mathbf{k}_2^* \otimes \mathbf{k}_1^*],
\end{aligned}
$$

where $\mathbf{U}_{\theta^{(s)}}$ is the $k_1k_2 \times (k_1k_2 - 4)$ matrix whose columns are eigenvectors of $\mathbf{P}_s$ having non null-eigenvalues, $\mathbf{U}_{\theta^{(t)}}$ is the $k_t \times (k_t - 2)$ matrix whose columns are eigenvectors of $\mathbf{P}_t$ having non null-eigenvalues and $\mathbf{U}_{\theta^{(st)}}$ is the $k_1k_2k_t \times (k_1k_2k_t - 8)$ matrix whose columns are eigenvectors of $\mathbf{P}_{st}$ having non null-eigenvalues. In this case, reparameterizing the random effects as described above and removing the repeated columns of the design matrices, leads to the model

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

$$\begin{aligned}\log \mathbf{r} = {} & (\mathbf{1}_T \otimes \mathbf{1}_n)\eta + [\mathbf{1}_T \otimes \mathbf{B}_s(\mathbf{1}_{k_2} \otimes \mathbf{k}_1^*) : \mathbf{1}_T \otimes \mathbf{B}_s(\mathbf{k}_2^* \otimes \mathbf{1}_{k_1}) : \mathbf{1}_T \otimes \mathbf{B}_s(\mathbf{k}_2^* \otimes \mathbf{k}_1^*)]\boldsymbol{\beta}_{\theta^{(s)}} \\ & + (\mathbf{1}_T \otimes \mathbf{B}_s\mathbf{U}_{\theta^{(s)}})\boldsymbol{\alpha}_{\theta^{(s)}} + (\mathbf{B}_t\mathbf{k}_t^* \otimes \mathbf{1}_n)\boldsymbol{\beta}_{\theta^{(t)}} + (\mathbf{B}_t\mathbf{U}_{\theta^{(t)}} \otimes \mathbf{1}_n)\boldsymbol{\alpha}_{\theta^{(t)}} \\ & + [\mathbf{B}_t\mathbf{k}_t^* \otimes \mathbf{B}_s(\mathbf{1}_{k_2} \otimes \mathbf{k}_1^*) : \mathbf{B}_t\mathbf{k}_t^* \otimes \mathbf{B}_s(\mathbf{k}_2^* \otimes \mathbf{1}_{k_1}) : \mathbf{B}_t\mathbf{k}_t^* \otimes \mathbf{B}_s(\mathbf{k}_2^* \otimes \mathbf{k}_1^*)]\boldsymbol{\beta}_{\theta^{(st)}} \\ & + [(\mathbf{B}_t \otimes \mathbf{B}_s)\mathbf{U}_{\theta^{(st)}}]\boldsymbol{\alpha}_{\theta^{(st)}}.\end{aligned}$$

It can be shown that removing the repeated columns leads to the linear constraints

$$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)} = 0; \quad \sum_{k=1}^{k_t}\theta_{k}^{(t)} = 0; \quad \sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t}\theta_{ijk}^{(st)} = 0;$$

$$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t} t\theta_{ijk}^{(st)} = \sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t} i\theta_{ijk}^{(st)} = \sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t} j\theta_{ijk}^{(st)} = \sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t} ij\theta_{ijk}^{(st)} = 0.$$

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Appendix 4B: R code for model fitting in INLA

The R code to fit some of the one, two and three-dimensional Bayesian B-spline models described in Section 4.2 with INLA is detailed below. In each case, the data must be ordered according to the Kronecker product of the space-time interaction term; i.e., $f_i(\mathbf{x}_t)$ for one-dimensional B-spline models (Table 4.1), $f_t(\mathbf{x}_1, \mathbf{x}_2)$ for two-dimensional B-spline models (Table 4.3) and $f(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_t)$ for three-dimensional B-spline models (Section 2.2).

To built the marginal B-spline bases for longitude, latitude and time, the `spline.des()` function from the `splines` library has been used. In general, assuming that the degree of the B-spline basis is p and the number of internal intervals is q, the B-spline basis consists of $p+q$ B-splines (columns) for which a total of $2p+q+1$ knots are needed (see Eilers and Marx, 1996 for details). In this chapter, equidistant internal knots have been considered to construct the spatial (longitude and latitude) and temporal cubic ($p=3$) B-spline marginal bases.

One-dimensional B-spline models

The one-dimensional B-spline model described in Equation (4.1) can be expressed in matrix form as

$$\log \mathbf{r} = (\mathbf{1}_n \otimes \mathbf{1}_T)\eta + (\mathbf{I}_n \otimes \mathbf{1}_T)\boldsymbol{\xi} + (\mathbf{1}_n \otimes \mathbf{B}_t)\boldsymbol{\theta}^{(t)} + (\mathbf{I}_n \otimes \mathbf{B}_t)\boldsymbol{\theta}^{(st)} \qquad (4.5)$$

where $\mathbf{r} = (r_{11}, \ldots, r_{1T}, \ldots, r_{n1}, \ldots, r_{nT})'$, $\mathbf{1}_n$ and $\mathbf{1}_T$ are vectors of ones of length n and T respectively, $\mathbf{I}_n$ is a $n \times n$ identity matrix and $\mathbf{B}_t$ is the temporal B-spline basis of dimension $T \times k_t$ obtained from the time covariate $\mathbf{x}_t = (x_1, \ldots, x_T)'$. First, the temporal B-spline basis $\mathbf{B}_t$ is constructed

```
> p <- 3  ## Degree of the B-spline basis
>
> xt <- <time>
> xt <- (xt-min(xt))/(max(xt)-min(xt))
>
> dist <- (max(xt)-min(xt))/q
> xtl <- min(xt)-dist*0.05
> xtr <- max(xt)+dist*0.05
> dxt <- (xtr-xtl)/q
> knots <- seq(xtl-p*dxt, xtr+p*dxt, by=dxt)
>
> Bt <- spline.des(knots,xt,p+1)$design
> kt <- ncol(Bt)
```

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

where `time` is the vector containing the time points, `xt` is the time covariate scaled into the unit interval $[0, 1]$, `q` is the number of internal intervals, and `knots` is the vector containing the total number of knots needed to built the B-spline basis. For breast cancer mortality data analyzed in Section 4.3.1, $q = 5$ internal intervals have been considered, giving rise to a temporal cubic B-spline basis of dimension $T \times k_t = 21 \times 8$ (with $k_t = p + q$).

Then, the design matrices of the random effects given in Equation (4.5) are defined, as well as the data frame containing the variables of the model

```
> Ms <- kronecker(diag(n),matrix(1,t,1))
> B_t <- kronecker(matrix(1,n,1),Bt)
> B_st <- kronecker(diag(n),Bt)
>
> Data <- list(O=<observed>, E=<expected>,
>              intercept=c(1,rep(NA,n+kt+n*kt)),
>              ID.area=c(NA,1:n,rep(NA,kt+n*kt)),
>              ID.year=c(rep(NA,1+n),1:kt,rep(NA,n*kt)),
>              ID.area.year=c(rep(NA,1+n+kt),1:(n*kt)))
```

where `observed` and `expected` are the vectors of observed and expected cases respectively, and `n` and `t` are the number of areas and time periods for which data is available (`n=47` provinces and `t=21` years for breast cancer mortality data).

Next, the spatial neighborhood matrix $\mathbf{R}_\xi$ and the temporal structure matrix $\mathbf{P}_{k_t}$ are defined

```
> g <- inla.read.graph("prov_nb.inla")
> R.xi <- matrix(0, g$n, g$n)
> for (i in 1:g$n){
>   R.xi[i,i]=g$nnbs[[i]]
>   R.xi[i,g$nbs[[i]]]=-1
> }
> R.Leroux <- diag(n)-R.xi
>
> D1 <- diff(diag(kt),differences=1)
> P.RW1 <- t(D1)%*%D1
>
> D2 <- diff(diag(kt),differences=2)
> P.RW2 <- t(D2)%*%D2
```

where `"prov_nb.inla"` is an `inla.graph` object containing the neighboring structure of the Spanish provinces.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Finally, the `formula` object for the models described in Table 4.1 are defined below.

Type I interaction (with $\Delta_{d_t} \equiv 1_{st}$ order difference matrix)

```
> LC <- kronecker(matrix(1,1,n),matrix(1,1,t)%*%Bt)
>
> formula <- O ~ -1 + intercept +
>            f(ID.area, model="generic1", Cmatrix = R.Leroux, constr=TRUE,
>              hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
>            f(ID.year, model="rw1", constr=TRUE,
>              hyper=list(prec=list(prior=sdunif))) +
>            f(ID.area.year, model="iid", constr=FALSE,
>              hyper=list(prec=list(prior=sdunif)),
>              extraconstr=list(A=LC, e=0))
```

where the linear constraint over the interaction term is expressed as

$$\sum_{i=1}^{n}\sum_{t=1}^{T} f_i(x_t) = 0 \iff \sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbf{I}_n \otimes \mathbf{B}_t)\boldsymbol{\theta}^{(st)} = (\mathbf{1}_n' \otimes \mathbf{1}_T'\mathbf{B}_t)\boldsymbol{\theta}^{(st)} = 0.$$

Type I interaction (with $\Delta_{d_t} \equiv 2_{nd}$ order difference matrix)

```
> LC1 <- kronecker(matrix(1,1,n),matrix(1,1,t)%*%Bt)
> LC2 <- kronecker(matrix(1,1,n),matrix(1:t,1,t)%*%Bt)
>
> formula <- O ~ -1 + intercept +
>            f(ID.area, model="generic1", Cmatrix = R.Leroux, constr=TRUE,
>              hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
>            f(ID.year, model="rw2", constr=TRUE,
>              hyper=list(prec=list(prior=sdunif))) +
>            f(ID.area.year, model="iid", constr=FALSE,
>              hyper=list(prec=list(prior=sdunif)),
>              extraconstr=list(A=rbind(LC1,LC2), e=rep(0,2)))
```

where the linear constraints over the interaction term are expressed as

$$\sum_{i=1}^{n}\sum_{t=1}^{T} f_i(x_t) = 0 \iff \sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbf{I}_n \otimes \mathbf{B}_t)\boldsymbol{\theta}^{(st)} = (\mathbf{1}_n' \otimes \mathbf{1}_T'\mathbf{B}_t)\boldsymbol{\theta}^{(st)} = 0,$$

$$\sum_{i=1}^{n}\sum_{t=1}^{T} t f_i(x_t) = 0 \iff \sum_{i=1}^{n}\sum_{t=1}^{T} t(\mathbf{I}_n \otimes \mathbf{B}_t)\boldsymbol{\theta}^{(st)} = (\mathbf{1}_n' \otimes \mathbf{t}^{*'}\mathbf{B}_t)\boldsymbol{\theta}^{(st)} = 0,$$

with $\mathbf{t}^* = (1, 2, \ldots, T)'$.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Type II interaction (with $\Delta_{d_t} \equiv 1_{st}$ order difference matrix)

```
> R <- kronecker(diag(n),P.RW1)
> r.def <- n
> A.constr.ST <- kronecker(diag(n),matrix(1,1,kt))
>
> formula <- O ~ -1 + intercept +
>            f(ID.area, model="generic1", Cmatrix = R.Leroux, constr=TRUE,
>              hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
>            f(ID.year, model="rw1", constr=TRUE,
>              hyper=list(prec=list(prior=sdunif))) +
>            f(ID.area.year, model="generic0", Cmatrix=R, rankdef=r.def,
>              constr=TRUE, hyper=list(prec=list(prior=sdunif)),
>              extraconstr=list(A=A.constr.ST, e=rep(0,n)))
```

where the linear constraints over the interaction term are expressed as

$$\sum_{k=1}^{k_t} \theta_{ik}^{(st)} = 0, \text{ for } i = 1, \ldots, n \iff (\mathbf{I}_n \otimes \mathbf{1}_{k_t}^{'})\boldsymbol{\theta}^{(st)} = \mathbf{0}.$$

Type II interaction (with $\Delta_{d_t} \equiv 2_{nd}$ order difference matrix)

```
> R <- kronecker(diag(n),P.RW2)
> r.def <- 2*n
> A.constr.ST <- kronecker(diag(n),matrix(1,1,kt))
>
> formula <- O ~ -1 + intercept +
>            f(ID.area, model="generic1", Cmatrix = R.Leroux, constr=TRUE,
>              hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
>            f(ID.year, model="rw2", constr=TRUE,
>              hyper=list(prec=list(prior=sdunif)),
>              extraconstr=list(A=matrix(1:kt,1,kt),e=0)) +
>            f(ID.area.year, model="generic0", Cmatrix=R, rankdef=r.def,
>              constr=TRUE, hyper=list(prec=list(prior=sdunif)),
>              extraconstr=list(A=A.constr.ST, e=rep(0,n)))
```

where the linear constraints over the temporal and interaction term are expressed as

$$\sum_{k=1}^{k_t} \theta_k^{(t)} = \sum_{k=1}^{k_t} k\theta_k^{(t)} = 0 \iff \begin{bmatrix} \mathbf{1}_{k_t}^{'} \\ \mathbf{k}_t^{*'} \end{bmatrix} \boldsymbol{\theta}^{(t)} = \mathbf{0},$$

$$\sum_{k=1}^{k_t} \theta_{ik}^{(st)} = 0, \text{ for } i = 1, \ldots, n \iff (\mathbf{I}_n \otimes \mathbf{1}_{k_t}^{'})\boldsymbol{\theta}^{(st)} = \mathbf{0}.$$

with $\mathbf{k}_t^* = (1, 2, \ldots, k_t)^{'}$.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Type III interaction (with $\Delta_{d_t} \equiv 1_{st}$ order difference matrix)

```
> R <- kronecker(R.xi,diag(kt))
> r.def <- kt
> A.constr.ST <- kronecker(matrix(1,1,n),diag(kt))
>
> formula <- O ~ -1 + intercept +
>            f(ID.area, model="generic1", Cmatrix = R.Leroux, constr=TRUE,
>              hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
>            f(ID.year, model="rw1", constr=TRUE,
>              hyper=list(prec=list(prior=sdunif))) +
>            f(ID.area.year, model="generic0", Cmatrix=R, rankdef=r.def,
>              constr=TRUE, hyper=list(prec=list(prior=sdunif)),
>              extraconstr=list(A=A.constr.ST, e=rep(0,kt)))
```

where the linear constraints over the interaction term are expressed as

$$\sum_{i=1}^{n} \theta_{ik}^{(st)} = 0, \text{ for } k = 1, \ldots, k_t \iff (\mathbf{1}_n^{'} \otimes \mathbf{I}_{k_t})\boldsymbol{\theta}^{(st)} = \mathbf{0}.$$

Type III interaction (with $\Delta_{d_t} \equiv 2_{nd}$ order difference matrix)

```
> R <- kronecker(R.xi,diag(kt))
> r.def <- kt
> A.constr.ST <- kronecker(matrix(1,1,n),diag(kt))
>
> formula <- O ~ -1 + intercept +
>            f(ID.area, model="generic1", Cmatrix = R.Leroux, constr=TRUE,
>              hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
>            f(ID.year, model="rw2", constr=TRUE,
>              hyper=list(prec=list(prior=sdunif))) +
>            f(ID.area.year, model="generic0", Cmatrix=R, rankdef=r.def,
>              constr=TRUE, hyper=list(prec=list(prior=sdunif)),
>              extraconstr=list(A=A.constr.ST, e=rep(0,kt)))
```

The same constraints as in the previous model (Type III interaction with $\Delta_{d_t} \equiv 1_{st}$ order penalty) are considered here.

Type IV interaction (with $\Delta_{d_t} \equiv 1_{st}$ order difference matrix)

```
> R <- kronecker(R.xi,P.RW1)
> r.def <- n+kt-1
> A1 <- kronecker(diag(n),matrix(1,1,kt))
> A2 <- kronecker(matrix(1,1,n),diag(kt))
```

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

```
> A.constr.ST <- rbind(A1,A2)
>
> formula <- O ~ -1 + intercept +
>            f(ID.area, model="generic1", Cmatrix = R.Leroux, constr=TRUE,
>              hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
>            f(ID.year, model="rw1", constr=TRUE,
>              hyper=list(prec=list(prior=sdunif))) +
>            f(ID.area.year, model="generic0", Cmatrix=R, rankdef=r.def,
>              constr=TRUE, hyper=list(prec=list(prior=sdunif)),
>              extraconstr=list(A=A.constr.ST, e=rep(0,n+kt)))
```

where the linear constraints over the interaction term are expressed as

$$\sum_{k=1}^{k_t} \theta_{ik}^{(st)} = 0, \text{ for } i = 1, \ldots, n \iff (\mathbf{I}_n \otimes \mathbf{1}'_{k_t})\boldsymbol{\theta}^{(st)} = \mathbf{0}.$$

$$\sum_{i=1}^{n} \theta_{ik}^{(st)} = 0, \text{ for } k = 1, \ldots, k_t \iff (\mathbf{1}'_{n} \otimes \mathbf{I}_{k_t})\boldsymbol{\theta}^{(st)} = \mathbf{0}.$$

Type IV interaction (with $\Delta_{d_t} \equiv 2_{nd}$ order difference matrix)

```
> R <- kronecker(R.xi,P.RW2)
> r.def <- n+2*kt-1
> A1 <- kronecker(diag(n),matrix(1,1,kt))
> A2 <- kronecker(matrix(1,1,n),diag(kt))
> A.constr.ST <- rbind(A1,A2)
>
> formula <- O ~ -1 + intercept +
>            f(ID.area, model="generic1", Cmatrix = R.Leroux, constr=TRUE,
>              hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
>            f(ID.year, model="rw2", constr=TRUE,
>              hyper=list(prec=list(prior=sdunif))) +
>            f(ID.area.year, model="generic0", Cmatrix=R, rankdef=r.def,
>              constr=TRUE, hyper=list(prec=list(prior=sdunif)),
>              extraconstr=list(A=A.constr.ST, e=rep(0,n+kt)))
```

The same constraints as in the previous model (Type IV interaction with $\Delta_{d_t} \equiv 1_{st}$ order penalty) are considered here.

Finally, we run the INLA algorithm with a call to the `inla()` function as

```
> inla(formula, family="poisson", data=Data, E=E,
>      control.predictor=list(A=cBind(rep(1,n*t), Ms, B_t, B_st),
>                             compute=TRUE, link=1, cdf=c(log(1))),
```

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

```
>       control.compute=list(dic=TRUE, cpo=TRUE, waic=TRUE),
>       control.inla=list(strategy="laplace"))
```

Note that the linear predictor is modeled as a linear combination of the random effects defined by the design matrices described above through the `control.predictor` argument (see Section 1.4.4).

Two-dimensional B-spline models

The two-dimensional B-spline model described in Equation (4.2) can be expressed in matrix form as

$$\log \mathbf{r} = (\mathbf{1}_T \otimes \mathbf{1}_n)\eta + (\mathbf{1}_T \otimes \mathbf{B}_s)\boldsymbol{\theta}^{(s)} + (\mathbf{I}_T \otimes \mathbf{1}_n)\boldsymbol{\gamma} + (\mathbf{I}_T \otimes \mathbf{B}_s)\boldsymbol{\theta}^{(st)} \tag{4.6}$$

where $\mathbf{r} = (r_{11}, \ldots, r_{n1}, \ldots, r_{1T}, \ldots, r_{nT})'$, $\mathbf{1}_n$ and $\mathbf{1}_T$ are vectors of ones of length n and T respectively, $\mathbf{I}_T$ is a $T \times T$ identity matrix and $\mathbf{B}_s$ is the spatial two-dimensional B-spline basis of dimension $n \times k_1 k_2$ obtained from the row-wise Kronecker product of the marginal bases for longitude $\mathbf{x}_1 = (x_{11}, \ldots, x_{1n})'$ and latitude $\mathbf{x}_2 = (x_{21}, \ldots, x_{2n})'$.

First, the spatial two-dimensional B-spline basis $\mathbf{B}_s$ is constructed from the marginal bases of longitude $\mathbf{B}_1$ and latitude $\mathbf{B}_2$

```
> p <- 3  ## Degree of the B-spline basis
>
> ## Marginal basis for longitude ##
> x1 <- <longitude>
> x1 <- (x1-min(x1))/(max(x1)-min(x1))
>
> dist1 <- (max(x1)-min(x1))/q
> x1l <- min(x1)-dist1*0.05
> x1r <- max(x1)+dist1*0.05
> dx1 <- (x1r-x1l)/q
> knots1 <- seq(x1l-p*dx1, x1r+p*dx1, by=dx1)
>
> B1 <- spline.des(knots1,x1,p+1)$design
> k1 <- ncol(B1)
>
> ## Marginal basis for latitude ##
> x2 <- <latitude>
> x2 <- (x2-min(x2))/(max(x2)-min(x2))
>
> dist2 <- (max(x2)-min(x2))/q
```

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

```
> x2l <- min(x2)-dist2*0.05
> x2r <- max(x2)+dist2*0.05
> dx2 <- (x2r-x2l)/q
> knots2 <- seq(x2l-p*dx2, x2r+p*dx2, by=dx2)
>
> B2 <- spline.des(knots2,x2,p+1)$design
> k2 <- ncol(B2)
>
> ## Row-wise Kronecker product ##
> Rten <- function(X1,X2){
>   one1 <- matrix(1,1,ncol(X1))
>   one2 <- matrix(1,1,ncol(X2))
>   kronecker(X1,one2)*kronecker(one1,X2)
> }
>
> Bs <- Rten(B2,B1)
> ks <- ncol(Bs)
```

where `longitude` and `latitude` are the vectors containing the coordinates of the centroids of the areas, `x1` and `x2` are the longitude and latitude covariates scaled into the unit intervals $[0, 1]$, `q` is the number of internal intervals, and `knots1` and `knots2` are the vectors containing the total number of knots needed to built the marginal B-spline basis $\mathbf{B}_1$ and $\mathbf{B}_2$. For breast cancer mortality data analyzed in Section 4.3.1, $q = 7$ internal intervals have been considered (for both longitude and latitude), giving rise to a spatial cubic B-spline basis of dimension $n \times k_1 k_2 = 47 \times 100$ (with $k_1 = k_2 = p + q$).

As for the previous models, the design matrices of the random effects given in Equation (4.6) are defined, as well as the data frame containing the variables of the model

```
> B_s <- kronecker(matrix(1,t,1),Bs)
> Mt <- kronecker(diag(t),matrix(1,n,1))
> B_st <- kronecker(diag(t),Bs)
>
> Data <- list(O=<observed>, E=<expected>,
>              intercept=c(1,rep(NA,ks+t+ks*t)),
>              ID.area=c(NA,1:ks,rep(NA,t+ks*t)),
>              ID.year=c(rep(NA,1+ks),1:t,rep(NA,ks*t)),
>              ID.area.year=c(rep(NA,1+ks+t),1:(ks*t)))
```

Then, the spatial structure matrix $\mathbf{P}_s$ and the temporal structure matrix $\mathbf{R}_\gamma$ are defined. The `R` code to fit the anisotropic precision matrix (using the `generic3` model) defined in Equation (4.3) is shown below

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

```
> ## Spatial structure matrix ##
> D1 <- diff(diag(k1),differences=<order>)
> P1 <- t(D1)%*%D1
>
> D2 <- diff(diag(k2),differences=<order>)
> P2 <- t(D2)%*%D2
>
> R1 <- kronecker(diag(k2),P1)
> R2 <- kronecker(P2,diag(k1))
>
> Cmat.s <- list(inla.as.sparse(R1),inla.as.sparse(R2))
>
> ## Temporal structure matrix ##
> D1 <- diff(diag(t),differences=1)
> P.RW1 <- t(D1)%*%D1
>
> D2 <- diff(diag(t),differences=2)
> P.RW2 <- t(D2)%*%D2
```

where D1 and D2 are the difference matrices Δ_{d_1} and Δ_{d_2} respectively. As in the analysis of breast cancer mortality data, second order difference matrices will be considered here for longitude and latitude (order=2).

The formula object for the models described in Table 4.3 are defined below.

Type I interaction (with $\Delta_{d_t} \equiv 1_{st}$ order difference matrix)

```
> LC1 <- kronecker(matrix(1,1,t),matrix(1,1,n)%*%Bs)
> LC2 <- kronecker(matrix(1,1,t),x1%*%Bs)
> LC3 <- kronecker(matrix(1,1,t),x2%*%Bs)
> LC4 <- kronecker(matrix(1,1,t),(x1*x2)%*%Bs)
>
> formula <- O ~ -1 + intercept +
>            f(ID.area, model="generic3", Cmatrix=Cmat.s, constr=TRUE,
>              hyper=list(prec1=list(prior=sdunif),prec2=list(prior=sdunif))) +
>            f(ID.year, model="rw1", constr=TRUE,
>              hyper=list(prec=list(prior=sdunif))) +
>            f(ID.area.year, model="iid", constr=FALSE,
>              hyper=list(prec=list(prior=sdunif)),
>              extraconstr=list(A=rbind(LC1,LC2,LC3,LC4),e=rep(0,4)))
```

where the linear constraints over the interaction term are expressed as

$$\sum_{i=1}^{n}\sum_{t=1}^{T} f_t(x_{1i}, x_{2i}) = 0 \iff \sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbf{I}_T \otimes \mathbf{B}_s)\boldsymbol{\theta}^{(st)} = (\mathbf{1}_T' \otimes \mathbf{1}_n'\mathbf{B}_s)\boldsymbol{\theta}^{(st)} = 0,$$

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

$$\sum_{i=1}^{n}\sum_{t=1}^{T} x_{2i} f_t(x_{1i}, x_{2i}) = 0 \iff \sum_{i=1}^{n}\sum_{t=1}^{T} x_{2i}(\mathbf{I}_T \otimes \mathbf{B}_s)\boldsymbol{\theta}^{(st)} = 0$$
$$\iff (\mathbf{1}_T' \otimes \mathbf{x}_2'\mathbf{B}_s)\boldsymbol{\theta}^{(st)} = 0,$$

$$\sum_{i=1}^{n}\sum_{t=1}^{T} x_{1i}x_{2i} f_t(x_{1i}, x_{2i}) = 0 \iff \sum_{i=1}^{n}\sum_{t=1}^{T} x_{1i}x_{2i}(\mathbf{I}_T \otimes \mathbf{B}_s)\boldsymbol{\theta}^{(st)} = 0$$
$$\iff (\mathbf{1}_T' \otimes \mathbf{x}_1'\mathbf{x}_2'\mathbf{B}_s)\boldsymbol{\theta}^{(st)} = 0.$$

Type I interaction (with $\Delta_{d_t} \equiv 2_{nd}$ order difference matrix)

```
> LC1 <- kronecker(matrix(1,1,t),matrix(1,1,n)%*%Bs)
> LC2 <- kronecker(matrix(1,1,t),x1%*%Bs)
> LC3 <- kronecker(matrix(1,1,t),x2%*%Bs)
> LC4 <- kronecker(matrix(1,1,t),(x1*x2)%*%Bs)
> LC5 <- kronecker(matrix(1:t,1,t),matrix(1,1,n)%*%Bs)
>
> formula <- O ~ -1 + intercept +
>            f(ID.area, model="generic3", Cmatrix=Cmat.s, constr=TRUE,
>              hyper=list(prec1=list(prior=sdunif),prec2=list(prior=sdunif))) +
>            f(ID.year, model="rw2", constr=TRUE,
>              hyper=list(prec=list(prior=sdunif))) +
>            f(ID.area.year, model="iid", constr=FALSE,
>              hyper=list(prec=list(prior=sdunif)),
>              extraconstr=list(A=rbind(LC1,LC2,LC3,LC4,LC5),e=rep(0,5)))
```

where besides the constraints considered for the previous model, the following one is also imposed over the interaction term

$$\sum_{i=1}^{n}\sum_{t=1}^{T} t f_t(x_{1i}, x_{2i}) = 0 \iff \sum_{i=1}^{n}\sum_{t=1}^{T} t(\mathbf{I}_t \otimes \mathbf{B}_s)\boldsymbol{\theta}^{(st)} = (\mathbf{t}^{*'} \otimes \mathbf{1}_n'\mathbf{B}_s)\boldsymbol{\theta}^{(st)} = 0.$$

Type II interaction (with $\Delta_{d_t} \equiv 1_{st}$ order difference matrix)

```
> R <- kronecker(P.RW1,diag(ks))
> r.def <- ks
> A.constr.ST <- kronecker(matrix(1,1,t),diag(ks))
>
> formula <- O ~ -1 + intercept +
>            f(ID.area, model="generic3", Cmatrix=Cmat.s, constr=TRUE,
>              hyper=list(prec1=list(prior=sdunif),prec2=list(prior=sdunif))) +
>            f(ID.year, model="rw1", constr=TRUE,
>              hyper=list(prec=list(prior=sdunif))) +
>            f(ID.area.year, model="generic0", Cmatrix=R, rankdef=r.def,
>              constr=TRUE, hyper=list(prec=list(prior=sdunif)),
>              extraconstr=list(A=A.constr.ST, e=rep(0,ks)))
```

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

where the linear constraints over the interaction term are expressed as

$$\sum_{t=1}^{T} \theta_{ij,t}^{(st)} = 0, \text{ for } \begin{matrix} i = 1, \ldots, k_1 \\ j = 1, \ldots, k_2 \end{matrix} \iff (\mathbf{1}_T^{'} \otimes \mathbf{I}_{k_1 k_2})\boldsymbol{\theta}^{(st)} = \mathbf{0}.$$

Type II interaction (with $\Delta_{d_t} \equiv 2_{nd}$ order difference matrix)

```
> R <- kronecker(P.RW2,diag(ks))
> r.def <- 2*ks
> A.constr.ST <- kronecker(matrix(1,1,t),diag(ks))
>
> formula <- O ~ -1 + intercept +
>            f(ID.area, model="generic3", Cmatrix=Cmat.s, constr=TRUE,
>              hyper=list(prec1=list(prior=sdunif),prec2=list(prior=sdunif))) +
>            f(ID.year, model="rw2", constr=TRUE,
>              hyper=list(prec=list(prior=sdunif)),
>              extraconstr=list(A=matrix(1:t,1,t),e=0)) +
>            f(ID.area.year, model="generic0", Cmatrix=R, rankdef=r.def,
>              constr=TRUE, hyper=list(prec=list(prior=sdunif)),
>              extraconstr=list(A=A.constr.ST, e=rep(0,ks)))
```

where the linear constraints over the temporal and the interaction term are expressed as

$$\sum_{t=1}^{T} \gamma_t = \sum_{t=1}^{T} t\gamma_t = 0 \iff \begin{bmatrix} \mathbf{1}_T^{'} \\ \mathbf{t}^{*'} \end{bmatrix} \boldsymbol{\gamma} = \mathbf{0},$$

$$\sum_{t=1}^{T} \theta_{ij,t}^{(st)} = 0, \text{ for } \begin{matrix} i = 1, \ldots, k_1 \\ j = 1, \ldots, k_2 \end{matrix} \iff (\mathbf{1}_T^{'} \otimes \mathbf{I}_{k_1 k_2})\boldsymbol{\theta}^{(st)} = \mathbf{0}.$$

Type III interaction (with $\Delta_{d_t} \equiv 1_{st}$ order difference matrix)

```
> ## Sum-to-zero constraints for the spatial term ##
> A1 <- kronecker(matrix(1,1,k2),matrix(1:k1,1,k1))
> A2 <- kronecker(matrix(1:k2,1,k2),matrix(1,1,k1))
> A3 <- kronecker(matrix(1:k2,1,k2),matrix(1:k1,1,k1))
> A.constr <- rbind(A1,A2,A3)
>
> ## Interaction structure matrix ##
> RR1 <- kronecker(diag(t),R1)
> RR2 <- kronecker(diag(t),R2)
> Cmat.st <- list(inla.as.sparse(RR1),inla.as.sparse(RR2))
>
> A.constr.ST <- kronecker(diag(t),matrix(1,1,ks))
```

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

```
>
> formula <- O ~ -1 + intercept +
>             f(ID.area, model="generic3", Cmatrix=Cmat.s, constr=TRUE,
>               hyper=list(prec1=list(prior=sdunif),prec2=list(prior=sdunif)),
>               extraconstr=list(A=A.constr, e=rep(0,3))) +
>             f(ID.year, model="rw1", constr=TRUE,
>               hyper=list(prec=list(prior=sdunif))) +
>             f(ID.area.year, model="generic3", Cmatrix=Cmat.st, constr=TRUE,
>               extraconstr=list(A=A.constr.ST, e=rep(0,t)),
>               hyper=list(prec1=list(prior=sdunif),prec2=list(prior=sdunif)))
```

where the linear constraints over the spatial and the interaction term are expressed as

$$\left.\begin{array}{l}\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij}^{(s)} = \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} i\theta_{ij}^{(s)} = 0 \\ \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} j\theta_{ij}^{(s)} = \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} ij\theta_{ij}^{(s)} = 0\end{array}\right. \iff \left[\begin{array}{c}(\mathbf{1}_{k_2}\otimes\mathbf{1}_{k_1})' \\ (\mathbf{1}_{k_2}\otimes\mathbf{k}_1^*)' \\ (\mathbf{k}_2^*\otimes\mathbf{1}_{k_1})' \\ (\mathbf{k}_2^*\otimes\mathbf{k}_1^*)'\end{array}\right]\boldsymbol{\theta}^{(s)} = \mathbf{0},$$

$$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\theta_{ij,t}^{(st)} = 0, \text{ for } t = 1,\ldots,T \iff (\mathbf{I}_T\otimes(\mathbf{1}_{k_2}\otimes\mathbf{1}_{k_1})')\boldsymbol{\theta}^{(st)} = \mathbf{0},$$

with $\mathbf{k}_1^* = (1,\ldots,k_1)'$ and $\mathbf{k}_2^* = (1,\ldots,k_2)'$.

Type III interaction (with $\Delta_{d_t} \equiv 2_{nd}$ order difference matrix)

```
> ## Sum-to-zero constraints for the spatial term ##
> A1 <- kronecker(matrix(1,1,k2),matrix(1:k1,1,k1))
> A2 <- kronecker(matrix(1:k2,1,k2),matrix(1,1,k1))
> A3 <- kronecker(matrix(1:k2,1,k2),matrix(1:k1,1,k1))
> A.constr <- rbind(A1,A2,A3)
>
> ## Interaction structure matrix ##
> RR1 <- kronecker(diag(t),R1)
> RR2 <- kronecker(diag(t),R2)
> Cmat.st <- list(inla.as.sparse(RR1),inla.as.sparse(RR2))
>
> A.constr.ST <- kronecker(diag(t),matrix(1,1,ks))
>
> formula <- O ~ -1 + intercept +
>             f(ID.area, model="generic3", Cmatrix=Cmat.s, constr=TRUE,
>               hyper=list(prec1=list(prior=sdunif),prec2=list(prior=sdunif)),
>               extraconstr=list(A=A.constr, e=rep(0,3))) +
>             f(ID.year, model="rw2", constr=TRUE,
>               hyper=list(prec=list(prior=sdunif))) +
```

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

```
>            f(ID.area.year, model="generic3", Cmatrix=Cmat.st, constr=TRUE,
>              extraconstr=list(A=A.constr.ST, e=rep(0,t)),
>              hyper=list(prec1=list(prior=sdunif),prec2=list(prior=sdunif)))
```

The same constraints as in the previous model (Type III interaction with $\Delta_{d_t} \equiv 1_{st}$ order penalty) are considered here.

Type IV interaction (with $\Delta_{d_t} \equiv 1_{st}$ order difference matrix)

```
> RR1 <- kronecker(P.RW1,R1)
> RR2 <- kronecker(P.RW1,R2)
> Cmat.st <- list(inla.as.sparse(RR1),inla.as.sparse(RR2))
>
> A1 <- kronecker(matrix(1,1,t),diag(ks))
> A2 <- kronecker(diag(t),matrix(1,1,ks))
> A.constr.ST <- rbind(A1,A2)
>
> formula <- O ~ -1 + intercept +
>            f(ID.area, model="generic3", Cmatrix=Cmat.s, constr=TRUE,
>              hyper=list(prec1=list(prior=sdunif),prec2=list(prior=sdunif))) +
>            f(ID.year, model="rw1", constr=TRUE,
>              hyper=list(prec=list(prior=sdunif))) +
>            f(ID.area.year, model="generic3", Cmatrix=Cmat.st, constr=TRUE,
>              extraconstr=list(A=A.constr.ST, e=rep(0,ks+t)),
>              hyper=list(prec1=list(prior=sdunif),prec2=list(prior=sdunif)))
```

where the linear constraints over the interaction term are expressed as

$$\sum_{t=1}^{T} \theta_{ij,t}^{(st)} = 0, \text{ for } \begin{matrix} i = 1, \ldots, k_1 \\ j = 1, \ldots, k_2 \end{matrix} \iff = (\mathbf{1}_T' \otimes \mathbf{I}_{k_1 k_2})\boldsymbol{\theta}^{(st)} = \mathbf{0},$$

$$\sum_{i=1}^{k_1} \sum_{j=1}^{k_2} \theta_{ij,t}^{(st)} = 0, \text{ for } t = 1, \ldots, T \iff (\mathbf{I}_T \otimes (\mathbf{1}_{k_2} \otimes \mathbf{1}_{k_1})')\boldsymbol{\theta}^{(st)} = \mathbf{0}.$$

Type IV interaction (with $\Delta_{d_t} \equiv 2_{nd}$ order difference matrix)

```
> RR1 <- kronecker(P.RW2,R1)
> RR2 <- kronecker(P.RW2,R2)
> Cmat.st <- list(inla.as.sparse(RR1),inla.as.sparse(RR2))
>
> A1 <- kronecker(matrix(1,1,t),diag(ks))
> A2 <- kronecker(diag(t),matrix(1,1,ks))
> A.constr.ST <- rbind(A1,A2)
>
> formula <- O ~ -1 + intercept +
>            f(ID.area, model="generic3", Cmatrix=Cmat.s, constr=TRUE,
```

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

```
>              hyper=list(prec1=list(prior=sdunif),prec2=list(prior=sdunif))) +
>            f(ID.year, model="rw2", constr=TRUE,
>              hyper=list(prec=list(prior=sdunif))) +
>            f(ID.area.year, model="generic3", Cmatrix=Cmat.st, constr=TRUE,
>              extraconstr=list(A=A.constr.ST, e=rep(0,ks+t)),
>              hyper=list(prec1=list(prior=sdunif),prec2=list(prior=sdunif)))
```

The same constraints as in the previous model (Type IV interaction with $\Delta_{d_t} \equiv 1_{st}$ order penalty) are considered here.

Finally, we run the INLA algorithm with a call to the `inla()` function as

```
> inla(formula, family="poisson", data=Data, E=E,
>      control.predictor=list(A=cBind(rep(1,n*t), B_s, Mt, B_st),
>                             compute=TRUE, link=1, cdf=c(log(1))),
>      control.compute=list(dic=TRUE, cpo=TRUE, waic=TRUE),
>      control.inla=list(strategy="laplace"))
```

Again, the linear predictor is modeled as a linear combination of the random effects defined by the design matrices described above.

Three-dimensional B-spline models

The ANOVA-type P-spline model described in Equation (2.3) can be expressed in matrix form as

$$\log \mathbf{r} = (\mathbf{1}_T \otimes \mathbf{1}_n)\eta + (\mathbf{1}_T \otimes \mathbf{B}_s)\boldsymbol{\theta}^{(s)} + (\mathbf{B}_t \otimes \mathbf{1}_n)\boldsymbol{\theta}^{(t)} + (\mathbf{B}_t \otimes \mathbf{B}_s)\boldsymbol{\theta}^{(st)} \tag{4.7}$$

where $\mathbf{r} = (r_{11}, \ldots, r_{n1}, \ldots, r_{1T}, \ldots, r_{nT})'$, $\mathbf{1}_n$ and $\mathbf{1}_T$ are vectors of ones of length n and T respectively, $\mathbf{B}_t$ is the temporal B-spline basis of dimension $T \times k_t$ obtained from the time covariate $\mathbf{x}_t = (x_1, \ldots, x_T)'$ and $\mathbf{B}_s$ is the spatial two-dimensional B-spline basis of dimension $n \times k_1 k_2$ obtained from the row-wise Kronecker product of the marginal bases for longitude $\mathbf{x}_1 = (x_{11}, \ldots, x_{1n})'$ and latitude $\mathbf{x}_2 = (x_{21}, \ldots, x_{2n})'$.

The same marginal B-spline basis for longitude ($\mathbf{B}_1$), latitude ($\mathbf{B}_2$) and time ($\mathbf{B}_t$) covariates constructed for the one and two-dimensional models are considered here. First, the design matrices of the random effects given in Equation (4.7) and the data frame containing the variables of the models are defined

```
> B_s <- kronecker(matrix(1,t,1),Bs)
> B_t <- kronecker(Bt,matrix(1,n,1))
> B_st <- kronecker(Bt,Bs)
>
```

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

```
> Data <- list(O=<observed>, E=<expected>,
>              intercept=c(1,rep(NA,ks+kt+ks*kt)),
>              ID.area=c(NA,1:ks,rep(NA,kt+ks*kt)),
>              ID.year=c(rep(NA,1+ks),1:kt,rep(NA,ks*kt)),
>              ID.area.year=c(rep(NA,1+ks+kt),1:(ks*kt)))
```

Then, the structure matrices $\mathbf{P}_s$, $\mathbf{P}_{k_t}$ and $\mathbf{P}_{st}$ for the "generic3" model are constructed

```
> ## Spatial structure matrices ##
> D1 <- diff(diag(k1),differences=<order.S>)
> P1 <- t(D1)%*%D1
>
> D2 <- diff(diag(k2),differences=<order.S>)
> P2 <- t(D2)%*%D2
>
> R1 <- kronecker(diag(k2),P1)
> R2 <- kronecker(P2,diag(k1))
> Cmat.s <- list(inla.as.sparse(R1),inla.as.sparse(R2))
>
> ## Temporal structure matrix ##
> Dt <- diff(diag(kt),differences=<order.T>)
> Pt <- t(Dt)%*%Dt
> Cmat.t <- list(inla.as.sparse(Pt))
>
> ## Spatio-temporal structure matrices ##
> RR1 <- kronecker(diag(kt),kronecker(diag(k2),P1))
> RR2 <- kronecker(diag(kt),kronecker(P2,diag(k1)))
> RR3 <- kronecker(Pt,kronecker(diag(k2),diag(k1)))
> Cmat.st <- list(inla.as.sparse(RR1),
>                 inla.as.sparse(RR2),
>                 inla.as.sparse(RR3))
```

where `order.S` and `order.T` are the order (1 or 2) of the spatial and temporal difference matrices respectively. Finally, the `formula` object for the models described in Table 4.6 are defined below.

Model 1 (Δ_{d_1}, Δ_{d_2}, $\Delta_{d_t} \equiv 1_{st}$ order difference matrices)

```
> formula <- O ~ -1 + intercept +
>            f(ID.area, model="generic3", Cmatrix=Cmat.s, constr=TRUE,
>              hyper=list(prec1=list(prior=sdunif),prec2=list(prior=sdunif))) +
```

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

```
>            f(ID.year, model="generic3", Cmatrix=Cmat.t, constr=TRUE,
>              hyper=list(prec1=list(prior=sdunif))) +
>            f(ID.area.year, model="generic3", Cmatrix=Cmat.st, constr=TRUE,
>              hyper=list(prec1=list(prior=sdunif),
>                         prec2=list(prior=sdunif),
>                         prec3=list(prior=sdunif)))
```

where sum to zero constraints are imposed over the random effects through the `constr=TRUE` argument.

Model 2 (Δ_{d_1}, $\Delta_{d_2} \equiv 1_{st}$ and $\Delta_{d_t} \equiv 2_{nd}$ order difference matrices)

```
> LC1 <- kronecker(matrix(1:kt,1,kt),matrix(1,1,ks))
>
> formula <- O ~ -1 + intercept +
>            f(ID.area, model="generic3", Cmatrix=Cmat.s, constr=TRUE,
>              hyper=list(prec1=list(prior=sdunif),prec2=list(prior=sdunif))) +
>            f(ID.year, model="generic3", Cmatrix=Cmat.t, constr=TRUE,
>              hyper=list(prec1=list(prior=sdunif))) +
>            f(ID.area.year, model="generic3", Cmatrix=Cmat.st, constr=TRUE,
>              hyper=list(prec1=list(prior=sdunif),
>                         prec2=list(prior=sdunif),
>                         prec3=list(prior=sdunif)),
>              extraconstr=list(A=LC1, e=0))
```

where the linear constraints over the interaction term are expressed as

$$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t}\theta_{ijk}^{(st)} = \sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t} k\theta_{ijk}^{(st)} = 0 \iff \begin{bmatrix} \mathbf{1}_{k_t}^{'} \otimes \left(\mathbf{1}_{k_2} \otimes \mathbf{1}_{k_1}\right)' \\ \mathbf{k}_t^{*'} \otimes \left(\mathbf{1}_{k_2} \otimes \mathbf{1}_{k_1}\right)' \end{bmatrix} \boldsymbol{\theta}^{(st)} = \mathbf{0}.$$

Model 3 (Δ_{d_1}, $\Delta_{d_2} \equiv 2_{nd}$ and $\Delta_{d_t} \equiv 1_{st}$ order difference matrices)

```
> LC2 <- kronecker(matrix(1,1,kt),kronecker(matrix(1,1,k2),matrix(1:k1,1,k1)))
> LC3 <- kronecker(matrix(1,1,kt),kronecker(matrix(1:k2,1,k2),matrix(1,1,k1)))
> LC4 <- kronecker(matrix(1,1,kt),kronecker(matrix(1:k2,1,k2),matrix(1:k1,1,k1)))
>
> formula <- O ~ -1 + intercept +
>            f(ID.area, model="generic3", Cmatrix=Cmat.s, constr=TRUE,
>              hyper=list(prec1=list(prior=sdunif),prec2=list(prior=sdunif))) +
>            f(ID.year, model="generic3", Cmatrix=Cmat.t, constr=TRUE,
>              hyper=list(prec1=list(prior=sdunif))) +
>            f(ID.area.year, model="generic3", Cmatrix=Cmat.st, constr=TRUE,
>              hyper=list(prec1=list(prior=sdunif),
>                         prec2=list(prior=sdunif),
>                         prec3=list(prior=sdunif)),
>              extraconstr=list(A=rbind(LC2,LC3,LC4), e=rep(0,3)))
```

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

where the linear constraints over the interaction term are expressed as

$$\begin{matrix} \sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t}\theta_{ijk}^{(st)} = \sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t} i\theta_{ijk}^{(st)} = 0 \\ \sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t} j\theta_{ijk}^{(st)} = \sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\sum_{k=1}^{k_t} ij\theta_{ijk}^{(st)} = 0 \end{matrix} \iff \begin{bmatrix} \mathbf{1}_{k_t}' \otimes (\mathbf{1}_{k_2} \otimes \mathbf{1}_{k_1})' \\ \mathbf{1}_{k_t}' \otimes (\mathbf{1}_{k_2} \otimes \mathbf{k}_1^*)' \\ \mathbf{1}_{k_t}' \otimes (\mathbf{k}_2^* \otimes \mathbf{1}_{k_1})' \\ \mathbf{1}_{k_t}' \otimes (\mathbf{k}_2^* \otimes \mathbf{k}_1^*)' \end{bmatrix} \boldsymbol{\theta}^{(st)} = \mathbf{0}.$$

Model 4 ***(Δ_{d_1}, Δ_{d_2}, $\Delta_{d_t} \equiv 2_{nd}$ order difference matrices)***

```
> formula <- O ~ -1 + intercept +
>            f(ID.area, model="generic3", Cmatrix=Cmat.s, constr=TRUE,
>              hyper=list(prec1=list(prior=sdunif),prec2=list(prior=sdunif))) +
>            f(ID.year, model="generic3", Cmatrix=Cmat.t, constr=TRUE,
>              hyper=list(prec1=list(prior=sdunif))) +
>            f(ID.area.year, model="generic3", Cmatrix=Cmat.st, constr=TRUE,
>              hyper=list(prec1=list(prior=sdunif),
>                         prec2=list(prior=sdunif),
>                         prec3=list(prior=sdunif)),
>              extraconstr=list(A=rbind(LC1,LC2,LC3,LC4), e=rep(0,4)))
```

where the whole set of constraints is considered over the interaction term.

Finally, we run the INLA algorithm with a call to the `inla()` function as

```
> inla(formula, family="poisson", data=Data, E=E,
>      control.predictor=list(A=cBind(rep(1,n*t), B_s, B_t, B_st),
>                             compute=TRUE, link=1, cdf=c(log(1))),
>      control.compute=list(dic=TRUE, cpo=TRUE, waic=TRUE),
>      control.inla=list(strategy="laplace"))
```

modeling the linear predictor as a linear combination of the random effects defined by the design matrices described above.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Conclusions and further work

Despite the amount of work that has been published in the field of spatio-temporal disease mapping, there are still certain issues that need to be investigated. In Chapter 1, we provide a brief review of the literature in spatio-temporal disease mapping. The model proposed by Knorr-Held (2000) is described in detail. Throughout this dissertation, the integrated nested Laplace approximations (INLA) technique is used for model fitting and inference. The INLA methodology is briefly described and many useful details are provided facilitating its use within the `R-INLA` package.

Chapter 2 compares alternative space-time models in disease mapping analyzing the smoothing effects in space and time, and evaluating their ability to detect high-risks areas. An appropriate decomposition of the estimated log-risks is proposed defining comparable spatial, temporal, and spatio-temporal "patterns". By doing this, the smoothing effects in each dimension (space and time) can be compared in terms of bias and variability. We conclude that when the number of expected cases is very small (which is common when analyzing rare diseases or very small domains like municipalities or census tracks), P-spline models performs better in terms of high-risk area detection. Although smoother relative risk estimates are obtained using these models, higher values of true positive rates are observed in comparison with CAR-based models. However, some caution is recommended as convergence problems might occur.

In Chapter 3 spatio-temporal models including two-level of spatially structured random effects are proposed. These models permit both to identify regional effects and to model spatio-temporal interactions at different levels of spatial aggregation. Brain cancer mortality data are analyzed using these new model proposals aggregating municipalities within health areas, and differences are observed in the temporal trends of the health areas. Although models with two levels of spatial aggregation are only described, multilevel models can be also constructed if needed. However,

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

the computational burden for some of these models would increase significantly, especially for those where independent random effects are considered to model the spatial dependence within each SLA (named as two-level models E and F). The set of identifiability constraints are clearly established for all the model proposals. If SLA space-time interactions are considered instead of fitting FLA interactions, the number of constraints needed is much smaller, and consequently the computational time is highly reduced. The simulation study reveals that the proposed two-level models outperform the models described in Knorr-Held (2000) if different levels of spatial aggregation exist.

Finally, in Chapter 4, B-spline models are proposed to specify space-time interactions in Bayesian disease mapping. Some new model proposals are described, such as models including spatially correlated (or not) one-dimensional temporal B or P-splines, and models with temporally correlated (or not) two-dimensional B or P-splines. The models are compared with three-dimensional P-spline models within a fully Bayesian setting. Again, the necessary set of identifiability constraints are derived in detail. We observe that generally, the model with spatially correlated temporal P-splines (where similar temporal trends are estimated for neighboring regions) provides the best trade-off between model fit and complexity. However, this model may not be feasible from a computational point of view when the number of areas increases. In contrast, the three-dimensional P-spline models seem to be a promising alternative, obtaining accurate risk estimates in relatively shorter computational time.

Further work

In this dissertation, particular attention has been paid to solve identifiability issues in spatio-temporal disease mapping models by imposing the necessary sum-to-zero constraints on the random effects. Our models do not include explanatory variables, something which may be very interesting to assess potential risk factors. Identifiability issues in these models should be also investigated in detail, because ignoring the spatial or temporal correlation between the covariates and the random effects can lead to misleading results due to confounding issues.

A recent matter of research is multidimensional disease mapping in which several diseases are jointly analysed in space and time including also other factors such as age groups (see for example Martinez-Beneito, 2013; MacNab, 2016a,b; Martinez-Beneito et al., 2017 and the references therein). We would like to extend the spatio-temporal spline models described in Chapter 4 to a multidimensional framework, including several diseases and other variables of interest such us age or social deprivation. Reducing the computational burden inherent to these models and trying to implement them into an appropriate software are the biggest challenges.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Another future research target is to tackle simultaneously two contradictory goals in disease mapping: smoothing and clustering detection. Risks are generally smoothed to remove random noise and to unveil the underlying geographical and temporal patterns. However, an excess of smoothing may impede the identification of areas with extreme (high or low) risks as discontinuities is the smooth risk surface become blurred. We want to investigate new models for both clustering detection and smoothing risks in the presence of local discontinuities.

Finally, we would like to provide users with an easy-to-use tool to fit the spatio-temporal models described in this dissertation. In particular, we are currently working on developing a web tool application using Shiny (Chang et al., 2017). This is of particular interest for public health professionals.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

References

Adin, A., Martínez-Beneito, M., Botella-Rocamora, P., Goicoa, T., and Ugarte, M. (2017). Smoothing and high risk areas detection in space-time disease mapping: a comparison of P-splines, autoregressive, and moving average models. *Stochastic Environmental Research and Risk Assessment*, 31(2):403–415.

Ainsworth, L. and Dean, C. (2006). Approximate inference for disease mapping. *Computational Statistics & Data Analysis*, 50(10):2552–2570.

Assunção, R. M., Reis, I. A., and Oliveira, C. D. L. (2001). Diffusion and prediction of Leishmaniasis in a large metropolitan area in Brazil with a bayesian space–time model. *Statistics in Medicine*, 20(15):2319–2335.

Bauer, C., Wakefield, J., Rue, H., Self, S., Feng, Z., and Wang, Y. (2016). Bayesian penalized spline models for the analysis of spatio-temporal count data. *Statistics in Medicine*, 35(11):1848–1865.

Belitz, C. and Lang, S. (2008). Simultaneous selection of variables and smoothing parameters in structured additive regression models. *Computational Statistics & Data Analysis*, 53(1):61–81.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., and Songini, M. (1995). Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine*, 14(21-22):2433–2443.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20.

Best, N., Richardson, S., and Thomson, A. (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research*, 14(1):35–59.

Blangiardo, M. and Cameletti, M. (2015). *Spatial and Spatio-temporal Bayesian Models with R-INLA*. John Wiley & Sons.

Blangiardo, M., Cameletti, M., Baio, G., and Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-Temporal Epidemiology*, 7:39–55.

Botella-Rocamora, P. (2010). *Suavización espacio-temporal en cartografía de enfermedades*. PhD thesis, Universitat de València.

Botella-Rocamora, P., López-Quílez, A., and Martinez-Beneito, M. (2013). Spatial moving average risk smoothing. *Statistics in Medicine*, 32(15):2595–2612.

Breslow, N. (2004). Whither PQL? In *Proceedings of the Second Seattle Symposium in Biostatistics*, pages 1–22. Springer.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421):9–25.

Chang, W., Cheng, J., Allaire, J., Xie, Y., and McPherson, J. (2017). *shiny: Web Application Framework for R*. R package version 1.0.1.

Clayton, D. (1996). Generalized linear mixed models. In *Markov Chain Monte Carlo in Practice*, pages 275–301. Springer.

Czado, C., Gneiting, T., and Held, L. (2009). Predictive model assessment for count data. *Biometrics*, 65(4):1254–1261.

Dawid, A. P. (1984). Present position and potential developments: Some personal views: Statistical theory: The prequential approach. *Journal of the Royal Statistical Society. Series A (General)*, pages 278–292.

Dean, C., Ugarte, M. D., and Militino, A. F. (2004). Penalized quasi-likelihood with spatially correlated data. *Computational Statistics & Data Analysis*, 45(2):235–248.

Eilers, P. H., Currie, I. D., and Durbán, M. (2006). Fast and compact smoothing on large multidimensional grids. *Computational Statistics & Data Analysis*, 50(1):61–76.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, pages 89–102.

Etxeberria, J., Goicoa, T., Ugarte, M. D., and Militino, A. F. (2014). Evaluating space-time models for short-term cancer mortality risk predictions in small areas. *Biometrical Journal*, 56(3):383–402.

Etxeberria, J., Ugarte, M. D., Goicoa, T., and Militino, A. F. (2015). On predicting cancer mortality using ANOVA-type P-spline models. *REVSTAT–Statistical Journal*, 13(1):21–40.

Ferlay, J., Steliarova-Foucher, E., Lortet-Tieulent, J., Rosso, S., Coebergh, J., Comber, H., Forman, D., and Bray, F. (2013). Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *European Journal of Cancer*, 49(6):1374–1403.

Fong, Y., Rue, H., and Wakefield, J. (2010). Bayesian inference for generalized linear mixed models. *Biostatistics*, 11(3):397–412.

Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for bayesian models. *Statistics and Computing*, 24(6):997–1016.

Gelman, A. and Shirley, K. (2011). Inference from simulations and monitoring convergence. In Brooks, S., Gelman, A., Jones, G., and Meng, X., editors, *Handbook of Markov Chain Monte Carlo*, pages 163–174. CRC Press.

Gilks, W., Richardson, S., and Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Goicoa, T., Adin, A., Ugarte, M. D., and Hodges, J. (2017). In spatio-temporal disease mapping models, identifiability constraints affect PQL and INLA results. *Stochastic Environmental Research and Risk Assessment (online first, DOI:10.1007/s00477-017-1405-0)*.

Goicoa, T., Etxeberria, J., and Ugarte, M. D. (2016). Splines in disease mapping. In Lawson, A. B., Banerjee, S., Haining, R. P., and Ugarte, M. D., editors, *Handbook of Spatial Epidemiology*, chapter 12, pages 225–238. Chapman and Hall/CRC.

Goicoa, T., Ugarte, M. D., Etxeberria, J., and Militino, A. F. (2012). Comparing CAR and P-spline models in spatial disease mapping. *Environmental and Ecological Statistics*, 19(4):573–599.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Held, L., Schrödle, B., and Rue, H. (2010). Posterior and cross-validatory predictive checks: a comparison of MCMC and INLA. In *Statistical Modelling and Regression Structures*, pages 91–110. Springer.

Hodges, J. S. and Reich, B. J. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician*, 64(4):325–334.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, 19(17-18):2555–2567.

Knorr-Held, L. and Besag, J. (1998). Modelling risk from a disease in time and space. *Statistics in Medicine*, 17(18):2045–2060.

Knorr-Held, L. and Rue, H. (2002). On block updating in Markov random field models for disease mapping. *Scandinavian Journal of Statistics*, 29(4):597–614.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13(1):183–212.

Lee, D. (2011). A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spatial and Spatio-temporal Epidemiology*, 2(2):79–89.

Lee, D.-J. and Durbán, M. (2009). Smooth-CAR mixed models for spatial count data. *Computational Statistics & Data Analysis*, 53(8):2968–2979.

Leroux, B. G., Lei, X., and Breslow, N. (1999). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In Halloran, M. and Berry, D., editors, *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, pages 179–191. Springer-Verlag: New York.

López-Abente, G., Aragonés, N., Pérez-Gómez, B., Pollán, M., García-Pérez, J., Ramis, R., and Fernández-Navarro, P. (2014). Time trends in municipal distribution patterns of cancer mortality in Spain. *BMC Cancer*, 14(1):535.

Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*, 28(25):3049–3067.

MacNab, Y. and Dean, C. (2000). Parametric bootstrap and penalized quasi-likelihood inference in conditional autoregressive models. *Statistics in Medicine*, 19(17-18):2421–2435.

MacNab, Y. C. (2007). Spline smoothing in Bayesian disease mapping. *Environmetrics*, 18(7):727–744.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

MacNab, Y. C. (2011). On Gaussian Markov random fields and Bayesian disease mapping. *Statistical Methods in Medical Research*, 20(1):49–68.

MacNab, Y. C. (2014). On identification in Bayesian disease mapping and ecological–spatial regression models. *Statistical Methods in Medical Research*, 23(2):134–155.

MacNab, Y. C. (2016a). Linear models of coregionalization for multivariate lattice data: a general framework for coregionalized multivariate CAR models. *Statistics in Medicine*, 35(21):3827–3850.

MacNab, Y. C. (2016b). Linear models of coregionalization for multivariate lattice data: Order-dependent and order-free cMCARs. *Statistical Methods in Medical Research*, 25(4):1118–1144.

MacNab, Y. C. and Dean, C. (2001). Autoregressive spatial smoothing and temporal spline smoothing for mapping rates. *Biometrics*, 57(3):949–956.

MacNab, Y. C. and Gustafson, P. (2007). Regression B-spline smoothing in Bayesian disease mapping: with an application to patient safety surveillance. *Statistics in Medicine*, 26(24):4455–4474.

MacNab, Y. C. and Lin, Y. (2009). On empirical Bayes penalized quasi-likelihood inference in GLMMs and in Bayesian disease mapping and ecological modeling. *Computational Statistics & Data Analysis*, 53(8):2950–2967.

Marí-Dell'Olmo, M. and Martínez-Beneito, M. A. (2015). A Multilevel Regression Model for Geographical Studies in Sets of Non-Adjacent Cities. *PloS One*, 10(8):1–12.

Martinez-Beneito, M. A. (2013). A general modelling framework for multivariate disease mapping. *Biometrika*, 100(3):539–553.

Martinez-Beneito, M. A., Botella-Rocamora, P., Banerjee, S., et al. (2017). Towards a Multidimensional Approach to Bayesian Disease Mapping. *Bayesian Analysis*, 12(1):239–259.

Martínez-Beneito, M. A., López-Quilez, A., and Botella-Rocamora, P. (2008). An autoregressive approach to spatio-temporal disease mapping. *Statistics in Medicine*, 27(15):2874–2889.

Martino, S. and Rue, H. (2009). *Implementing Approximate Bayesian Inference using Integrated Nested Laplace Approximation: a manual for the* `INLA` *program.* Department of Mathematical Sciences, NTNT, Norway.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Martins, T. G., Simpson, D., Lindgren, F., and Rue, H. (2013). Bayesian computing with INLA: new features. *Computational Statistics & Data Analysis*, 67:68–83.

Pettit, L. (1990). The conditional predictive ordinate for the normal distribution. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 175–184.

Plummer, M. (2008). Penalized loss functions for Bayesian model comparison. *Biostatistics*, 9(3):523–539.

Plummer, M. (2016). *rjags: Bayesian Graphical Models using MCMC*. R package version 4-6.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reich, B. J., Hodges, J. S., and Zadnik, V. (2006). Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics*, 62(4):1197–1206.

Richardson, S., Thomson, A., Best, N., and Elliott, P. (2004). Interpreting posterior relative risk estimates in disease-mapping studies. *Environmental Health Perspectives*, pages 1016–1025.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, volume 104. Chapman & Hall/CRC.

Rue, H. and Martino, S. (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *Journal of Statistical Planning and Inference*, 137(10):3177–3192.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., and Lindgren, F. K. (2017). Bayesian computing with INLA: a review. *Annual Review of Statistics and Its Application*, 4:395–421.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric regression*. Number 12. Cambridge university press.

Schmid, V. and Held, L. (2004). Bayesian extrapolation of space–time trends in cancer registry data. *Biometrics*, 60(4):1034–1042.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Schrödle, B. and Held, L. (2011a). A primer on disease mapping and ecological regression using. *Computational Statistics*, 26(2):241–258.

Schrödle, B. and Held, L. (2011b). Spatio-temporal disease mapping using INLA. *Environmetrics*, 22(6):725–734.

Schrödle, B., Held, L., Riebler, A., and Danuser, J. (2011). Using integrated nested Laplace approximations for the evaluation of veterinary surveillance data from Switzerland: a case-study. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 60(2):261–279.

Simpson, D. P., Rue, H., Martins, T. G., Riebler, A., and Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science (in press)*.

Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003). *WinBUGS: version 1.4 User Manual*.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Stan Development Team (2016). *Stan Modeling Language User's Guide and Reference Manual*. Stan Version 2.14.0.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 44–47.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86.

Ugarte, M. D., Adin, A., and Goicoa, T. (2016). Two-level spatially structured models in spatio-temporal disease mapping. *Statistical Methods in Medical Research*, 25(4):1080–1100.

Ugarte, M. D., Adin, A., and Goicoa, T. (2017). One-dimensional, two-dimensional, and three dimensional B-splines to specify space-time interactions in Bayesian disease mapping: model fitting and model identifiability. *Spatial Statistics (accepted for publication, DOI:10.1016/j.spasta.2017.04.002)*.

Ugarte, M. D., Adin, A., Goicoa, T., Casado, I., Ardanaz, E., and Larrañaga, N. (2015a). Temporal evolution of brain cancer incidence in the municipalities of Navarre and the Basque Country, Spain. *BMC Public Health*, 15(1):1018.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Ugarte, M. D., Adin, A., Goicoa, T., and López-Abente, G. (2015b). Analyzing the evolution of young people's brain cancer mortality in Spanish provinces. *Cancer Epidemiology*, 39(3):480–485.

Ugarte, M. D., Adin, A., Goicoa, T., and Militino, A. F. (2014). On fitting spatio-temporal disease mapping models using approximate Bayesian inference. *Statistical Methods in Medical Research*, 23(6):507–530.

Ugarte, M. D., Goicoa, T., Etxeberria, J., and Militino, A. F. (2012a). A P-spline ANOVA type model in space-time disease mapping. *Stochastic Environmental Research and Risk Assessment*, 26(6):835–845.

Ugarte, M. D., Goicoa, T., Etxeberria, J., and Militino, A. F. (2012b). Projections of cancer mortality risks using spatio-temporal P-spline models. *Statistical Methods in Medical Research*, 21(5):545–560.

Ugarte, M. D., Goicoa, T., Etxeberria, J., Militino, A. F., and Pollán, M. (2010a). Age-specific spatio-temporal patterns of female breast cancer mortality in Spain (1975–2005). *Annals of Epidemiology*, 20(12):906–916.

Ugarte, M. D., Goicoa, T., Ibáñez, B., and Militino, A. F. (2009a). Evaluating the performance of spatio-temporal Bayesian models in disease mapping. *Environmetrics*, 20(6):647–665.

Ugarte, M. D., Goicoa, T., and Militino, A. F. (2009b). Empirical Bayes and Fully Bayes procedures to detect high-risk areas in disease mapping. *Computational Statistics & Data Analysis*, 53(8):2938–2949.

Ugarte, M. D., Goicoa, T., and Militino, A. F. (2010b). Spatio-temporal modeling of mortality risks using penalized splines. *Environmetrics*, 21(3-4):270–289.

Ugarte, M. D., Militino, A. F., and Goicoa, T. (2008). Prediction error estimators in empirical Bayes disease mapping. *Environmetrics*, 19(3):287–300.

Ventrucci, M. and Rue, H. (2016). Penalized complexity priors for degrees of freedom in Bayesian P-splines. *Statistical Modelling*, 16(6):429–453.

Wakefield, J. (2007). Disease mapping and spatial regression with count data. *Biostatistics*, 8(2):158–183.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(Dec):3571–3594.

Wood, S. (2006). *Generalized additive models: an introduction with R*. CRC press.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Wood, S. N., Scheipl, F., and Faraway, J. J. (2013). Straightforward intermediate rank tensor product smoothing in mixed models. *Statistics and Computing*, 23(3):341–360.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa